Validación Del Modelo DepressionDetect Para El Diagnóstico Automático De La Depresión Mediante La Voz Basado En Redes Convolucionales Profundas

INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

VALIDACIÓN DEL MODELO DEPRESSIONDETECT PARA EL DIAGNÓSTICO AUTOMÁTICO DE LA DEPRESIÓN MEDIANTE LA VOZ BASADO EN REDES CONVOLUCIONALES PROFUNDAS

EDWARD CAMILO VILLOTA TARAMUEL

UNIVERSIDAD TECNOLÓGICA DE PEREIRA

FACULTAD DE INGENIERÍAS

INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

PEREIRA – RISARALDA

2020

INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

VALIDACIÓN DEL MODELO DEPRESSIONDETECT PARA EL DIAGNÓSTICO AUTOMÁTICO DE LA DEPRESIÓN MEDIANTE LA VOZ BASADO EN REDES CONVOLUCIONALES PROFUNDAS

EDWARD CAMILO VILLOTA TARAMUEL

TRABAJO DE GRADO PARA OPTAR AL TÍTULO DE INGENIERO DE SISTEMAS Y COMPUTACIÓN

DIRECTOR DE PROYECTO

ANA MARÍA DE LAS MERCEDES LÓPEZ ECHEVERRY

UNIVERSIDAD TECNOLÓGICA DE PEREIRA

FACULTAD DE INGENIERÍAS

INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

PEREIRA – RISARALDA

2020

Nota de Aceptación:

_____

_____

_____

_____

_____

_____

Firma del Director de Proyecto

_____

Firma del Jurado

_____

Firma del Jurado

Pereira, 14 Agosto de 2020

DEDICATORIA

Dedico este proyecto de tesis a Dios y a mi familia.

A Dios quien me acompaña y ha estado conmigo a cada paso que doy, cuidándome y dándome fortaleza para continuar, a mi madre y mis abuelitos por ser un pilar importante en mi vida, quienes han velado por mi bienestar y educación siendo mi apoyo en todo momento.

Depositando su entera confianza en cada reto que se me presentaba sin dudar ni un solo momento en mi inteligencia y capacidad.

Es por ellos que soy lo que soy ahora. Los amo con mi vida esta meta es por ustedes y para el futuro que Dios nos depare.

AGRADECIMIENTOS

Mi más sinceros agradecimientos a la Directora Ana María López Echeverry y al Ingeniero Sebastián López Flórez por dirigir y asesorar respectivamente el proyecto de grado, por su colaboración, por su paciencia, por su compromiso, por el tiempo empleado, por los conocimientos brindados, por su constancia y orientación en la realización del presente trabajo de investigación, ya que supieron guiarme de la mejor manera posible con su repertorio amplio de conocimientos, obteniendo un crecimiento personal e intelectual, quedando sus enseñanzas plasmadas en mi memoria

Agradecimientos a todos y cada uno de los ingenieros de la facultad, porque de alguna manera supieron brindarme su gama de experiencia profesional, plasmadas en clases llenas de conocimiento

Agradezco también a todas las personas especiales para mí, por todo el apoyo brindado a lo largo de esta etapa de formación académica y personal que han contribuido de manera positiva en mi vida.

CONTENIDO

Pág.

INDICE DE TABLAS

ÍNDICE DE GRAFICAS

GLOSARIO

Anaconda: Es una plataforma gratuita y de código abierto escrita en python, orientada para ciencia de datos y aprendizaje automático, que brinda una gran cantidad de funcionalidades que permiten desarrollar aplicaciones de una manera más eficiente, rápida y sencilla [1].

Atención Primaria de Salud: Es la asistencia sanitaria esencial accesible a todos los individuos y familias de la comunidad a través de medios aceptables para ellos, con su plena participación y a un costo asequible para la comunidad y el país [2].

Autismo: Es un trastorno estático del desarrollo neurológico que persiste toda la vida y que incluye un amplio margen de alteraciones conductuales. De acuerdo al DSM-IV las manifestaciones clínicas distintivas son: sociabilidad alterada, anormalidades en el lenguaje y la comunicación no verbal, así como alteraciones en el margen de intereses y actividades. La deficiencia mental es frecuente, pero no universal. La perseveración, el aplanamiento afectivo y la falta de comprensión de los pensamientos y sentimientos de otros son notable [3].

CNN: Las redes neuronales convolucionales son un tipo de red neuronal artificial, en el que las conexiones de los nodos se inspiran en la corteza visual de los seres humanos. Las CNN son una herramienta poderosa en el reconocimiento de imágenes, el análisis de video y el procesamiento de lenguaje natural [4].

COVAREP: Es un repositorio de código abierto de algoritmos avanzados de procesamiento de voz [5].

CSV: (comma-separated values) es un archivo de texto que almacena los datos en forma de columnas, separadas por coma y las filas se distinguen por saltos de línea [6].

DAIC-WOZ: Base de datos que contiene entrevistas clínicas diseñadas para respaldar el diagnóstico de trastornos psicológicos como la ansiedad, la depresión y el estrés postraumático. Estas entrevistas se recopilaron como parte de un esfuerzo mayor para crear un agente informático que entreviste a personas e identifique indicadores verbales y no verbales de enfermedades mentales [7].

DataFrame: Son una clase de objetos especiales en el lenguaje de programación Python, denominados trama de datos, los cuales hacen uso de la librería pandas, el DataFrame permite almacenar y manipular datos tabulados en filas de observaciones y columnas de variables [8].

Demencia: Es un síndrome mayormente de naturaleza crónica o progresiva, causado por una variedad de enfermedades cerebrales que afectan la memoria, el

pensamiento, el comportamiento y la habilidad de realizar actividades de la vida diaria [9, p. 2].

Depresión: Los trastornos depresivos se caracterizan por un sentimiento de tristeza, pérdida de interés o de placer, sentimientos de culpa o autoestima baja, alteraciones del sueño o del apetito, fatiga y falta de concentración. La depresión puede ser duradera o recurrente, de modo que deteriora sustancialmente la capacidad de la persona de desempeñar su trabajo o rendir en sus estudios, o de hacer frente a su vida cotidiana. En su forma más severa, la depresión puede conducir al suicidio [10, p. 7].

DepressionDetect: Es un modelo automatizado para detectar la depresión de las características acústicas del habla, basado en redes convolucionales profundas CNN [4].

Discapacidad Intelectual: Es entendida como la adquisición lenta e incompleta de las habilidades cognitivas durante el desarrollo humano, que implica que la persona pueda tener dificultades para comprender, aprender y recordar cosas nuevas, que se manifiestan durante el desarrollo, y que contribuyen al nivel de inteligencia general, por ejemplo, habilidades cognitivas, motoras, sociales y de lenguaje [11, p. 2].

Discapacidad: Es un término genérico que abarca deficiencias, limitaciones de la actividad y restricciones a la participación. Se entiende por discapacidad la interacción entre las personas que padecen alguna enfermedad (por ejemplo, parálisis cerebral, síndrome de Down y depresión) y factores personales y ambientales (por ejemplo, actitudes negativas, transporte y edificios públicos inaccesibles y un apoyo social limitado) [12].

Espectrograma: Es una representación visual del sonido que muestra la amplitud de los componentes de frecuencia de una señal a lo largo del tiempo [4].

Esquizofrenia: Es una enfermedad grave que se inicia generalmente en la adolescencia tardía o en los primeros años de la edad adulta. Se caracteriza por distorsiones fundamentales de los procesos de pensamiento y percepción y por alteraciones de la afectividad. El trastorno afecta a las funciones más esenciales que confieren a las personas normales el sentimiento de individualidad, singularidad y autodirección. La firme creencia en ideas falsas y sin ninguna base real (delirios) es otra característica de este trastorno [13, p. 33].

Estrés postraumático: Este trastorno es una respuesta que una persona presenta después de haber estado presente en sucesos altamente estresantes. Algunos factores que causan estrés son la violencia sexual, los asaltos, los secuestros, el abuso sexual infantil, ser testigo presencial de una muerte, desastres naturales, guerras, accidentes automovilísticos, entre otros [14, p. 187].

Frecuencia de Muestreo: La frecuencia de muestreo (sample rate) indica el número de muestras por segundo que se toman de una señal de audio analógica (tiempo continuo) para transformarla en una señal de audio digital (tiempo discreto). [15].

Grupo etario: Según la Real Academia Española un grupo etario hace alusión a un grupo de varias personas que tienen la misma edad, o se encuentran en el mismo rango de edades.

Inteligencia Artificial: Es la ciencia de construir máquinas para que hagan cosas que, si las hicieran los humanos, requerirían inteligencia [16, p. 2].

Jupyter Notebook: o llamado Jupyter, es un entorno de trabajo interactivo que permite desarrollar código en Python de manera dinámica, a la vez que integrar en un mismo documento tanto bloques de código como texto, gráficas o imágenes. Es un SaaS utilizado ampliamente en análisis numérico, estadística y machine learning, entre otros campos de la informática y las matemáticas [17].

Librosa: Es un paquete de Python para análisis de audio y música. Proporciona los componentes básicos necesarios para crear sistemas de recuperación de información musical [18].

Likert: Escala de Likert es una herramienta de medición que, a diferencia de preguntas con respuesta sí/no, permite medir actitudes y conocer el grado de conformidad de un encuestado con cualquier afirmación que se le proponga, resultando especialmente útil emplearla en situaciones en las que desea que la persona matice su opinión [19].

Mago de Oz: Es un experimento de investigación en el que los sujetos interactúan con un sistema informático, que permite la recolección de diálogos persona-computador, simulando el computador con una persona oculta que realiza todas o algunas de las funciones que realizará el computador en el sistema de diálogo definitivo [20, p. 2].

Morbilidad: Según la Real Academia Española la morbilidad es la proporción de personas que enferman en un sitio y tiempo determinado.

Mortalidad: Según la Real Academia Española la mortalidad es la tasa de muertes producidas en una población durante un tiempo dado, en general o por una causa determinada.

Nyquist: Grupo de investigación perteneciente al programa de Ingeniería de Sistemas y Computación adscrito a la Facultad de Ingenierías de la Universidad Tecnológica de Pereira.

Observación subjetiva: Hace alusión según la Real Academia Española al modo de pensar o de sentir del sujeto, y no al objeto en sí mismo, en donde se realizan juicios de valor dejándose llevar por los sentimientos.

Pandas: Es una herramienta de manipulación de datos de alto nivel desarrollada por Wes McKinney. Es construido con el paquete Numpy y su estructura de datos clave es llamada DataFrame [21].

Patología: Según la Real Academia Española patología es el conjunto de síntomas de una enfermedad.

PHQ-8: El Cuestionario de salud del paciente de ocho ítems (PHQ) es un inventario de autoinforme de opción múltiple, que se utiliza como una herramienta de detección y diagnóstico para los trastornos de salud mental de la depresión, la ansiedad, el alcohol, la alimentación y los trastornos somáticos [22].

Psicofármacos: Son sustancias químicas que influyen en los procesos mentales actuando sobre el sistema nervioso.

Python: Es un lenguaje de programación interpretado que hace hincapié en la legibilidad del código, es un lenguaje de programación multiparadigma, soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional, python proporciona un equilibrio muy bueno entre lo práctico y lo conceptual, al ser un lenguaje interpretado, donde se puede tomar el lenguaje y empezar a hacer cosas interesantes casi inmediato, sin perderse en los problemas de compilación y enlazado [23, p. 6].

Salud Mental: Abarca una amplia gama de actividades directa o indirectamente relacionadas con el componente de bienestar mental incluido en la definición de salud que da la OMS: «un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades» [24].

Sesgo: Según la Real Academia Española el sesgo es un error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras.

TensorFlow: Es una plataforma de código abierto de extremo a extremo para el aprendizaje automático. Cuenta con un ecosistema integral y flexible de herramientas, bibliotecas y recursos de la comunidad que les permite a los investigadores impulsar un aprendizaje automático innovador y, a los desarrolladores, compilar e implementar con facilidad aplicaciones con tecnología de AA [25].

Trastorno afectivo bipolar: Consiste típicamente en episodios maníacos y depresivos interrumpidos por períodos en el que el estado de ánimo es normal. Los episodios maníacos manifiestan un estado de ánimo exaltado y de mayor energía,

lo que deriva en sobreactividad, habla atropellada o verborrea y menor necesidad de dormir [10, p. 7].

Trastorno mental: Los trastornos mentales y conductuales se consideran afecciones de importancia clínica, caracterizadas por alteraciones de los procesos de pensamiento, de la afectividad (emociones) o del comportamiento asociadas a angustia personal, a alteraciones del funcionamiento o a ambos. Para clasificarse como trastornos, estas anomalías deben ser duraderas o recurrentes, y deben causar cierta angustia personal o alteraciones del funcionamiento en una o más facetas de la vida [13, p. 21].

Trastornos somáticos: Se caracteriza por múltiples síntomas físicos persistentes que están asociados con pensamientos, sentimientos y comportamientos excesivos e inadaptados relacionados con esos síntomas [26].

WAV: WAVeform Audio Format, es un formato para almacenar sonido en archivos desarrollado en común por Microsoft e IBM [27].

RESUMEN

En este proyecto se plantea la validación del Modelo DepressionDetect para el diagnóstico automático de la depresión mediante la voz basado en redes convolucionales profundas por medio de un despliegue en condiciones controladas, con base en las características prosódicas del habla de una persona extraídas de la Base de Datos DAIC-WOZ, realizando un tratamiento a dichas señales de audio mediante métodos ya existentes; seguidamente procesando dicha información en la CNN a fin de obtener conclusiones que permitan detectar la depresión de una persona. De acuerdo con los resultados, el modelo sugerido es capaz de hacer predicciones con una precisión de 54%, referente a la clasificación de la depresión basado en las características prosódicas del habla de una persona.

1. INTRODUCCIÓN

Los problemas de Salud mental[1], como la depresión, la cual se caracteriza en gran medida como la principal causa de discapacidad y contribuye de forma muy importante a la carga mundial de morbilidad y mortalidad, lo que puede causar importantes pérdidas y cargas para los sistemas de salud, económico, social, educativo y de justicia.

Dado que la depresión es una condición común, incapacitante y que acorta la esperanza de vida de la persona que la padece [28, p. 2], es considerada una amenaza grave para la salud mental humana, donde esta puede afectar seriamente la vida normal de las personas, donde las personas con depresión pueden sentirse tristes, indefensas, vacías, ansiosas, anoréxicas, irritadas o molestas, y la depresión severa puede incluso conducir al suicidio [29, p. 2705], En Colombia, este trastorno mental afecta al 4,7 por ciento de la población [10, p. 18], ubicándose por encima del promedio mundial[2], lo que genera preocupación entre las autoridades en este tema.

Según la OMS [30], aunque hay tratamientos eficaces para la depresión, más de la mitad de los afectados en todo el mundo no reciben esos tratamientos. Uno de los principales obstáculos a una atención eficaz es la evaluación clínica errónea. En países de todo tipo de ingresos, las personas con depresión a menudo no son correctamente diagnosticadas, mientras que otras que en realidad no la padecen son a menudo diagnosticadas erróneamente y tratadas con antidepresivos. Los diagnósticos actuales son principalmente subjetivos, inconsistentes entre los profesionales y caros para la persona que puede necesitar ayuda [4, p. 1] además los primeros signos de depresión son difíciles de detectar y cuantificar, porque estos están limitados en gran medida por la observación subjetiva[3] de los médicos y la falta de un diagnóstico de seguimiento a largo plazo, ya que a medida que aumenta el número de pacientes con depresión, esto pone una carga extra a los médicos para diagnosticar con precisión el grado de depresión [29, p. 2705].

La tasa de detección de la depresión podría mejorar reduciendo su dependencia a la observación subjetiva de los médicos y proporcionando una evaluación mucho más objetiva y rápida, mediante la implementación de IA, varias investigaciones [31] [32] han demostrado que las señales del habla en los pacientes con depresión y en la gente sin esta condición tienen diferencias significativas; otros investigadores han

[1] La salud mental abarca una amplia gama de actividades directa o indirectamente relacionadas con el componente de bienestar mental incluido en la definición de salud que da la OMS: «Un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades» [24].

[2] Según la OMS a nivel mundial, se calcula que 4,4% de la población sufre un trastorno depresivo [10, p. 12].

[3] La observación subjetiva hace alusión según la Real Academia Española al modo de pensar o de sentir del sujeto, y no al objeto en sí mismo, en donde se realizan juicios de valor dejándose llevar por los sentimientos.

aplicado métodos [33] [34] de Deep Learning y [29] de Machine Learning para el análisis del audio para identificar el grado de depresión de pacientes con depresión.

Se busca con el presente proyecto reducir la carga de los médicos para diagnosticar una gran cantidad de síntomas depresivos, respaldar los diagnósticos de los profesionales médicos, al igual que promover la remisión, prevenir la recaída y reducir la carga emocional de la enfermedad en los pacientes [4, p. 1], mediante la detección automática de la depresión mediante redes neuronales convolucionales[4].

[4] Las redes neuronales convolucionales (CNN) son un tipo de red neuronal artificial, en el que las conexiones de los nodos se inspiran en la corteza visual de los seres humanos. Las CNN son una herramienta poderosa en el reconocimiento de imágenes, el análisis de video y el procesamiento de lenguaje natural [4]

2. PLANTEAMIENTO DEL PROBLEMA

2.1. Definición del Problema

2.1.1. Antecedentes del Problema

La salud mental está relacionada con la promoción del bienestar, la prevención de trastornos mentales y el tratamiento, rehabilitación de las personas afectadas por dichos trastornos. Entre estos trastornos se incluyen la depresión, estrés postraumático, la esquizofrenia, la demencia, las discapacidades intelectuales y los trastornos del desarrollo, como el autismo. De las anteriores mencionadas, la depresión es la principal causa mundial de discapacidad y contribuye de forma muy importante a la carga mundial general de morbilidad, mortalidad y discapacidad. Se calcula que en 2015 [10, p. 5] afectó a más de 300 millones de personas en el mundo (alrededor de 4,4 por ciento de la población total) aumentando en un 18 por ciento su alcance en la última década y que para el año 2020 ocupará el segundo lugar entre las causas más comunes de años de vida perdidos por discapacidad, y el primero en el 2030. En sus casos más graves, la depresión puede llevar al suicidio[5]. De acuerdo con la OMS [10, p. 14], cada año se suicidan cerca de 800 000 personas, y el suicidio es la segunda causa de muerte en el grupo etario de 15 a 29 años [10, p. 8]. En Colombia, este trastorno mental afecta al 4,7 por ciento de la población [10, p. 18], ubicándose por encima del promedio mundial, lo que genera preocupación entre las autoridades en este tema.

Según la OMS [30], aunque hay tratamientos eficaces para la depresión, más de la mitad de los afectados en todo el mundo (y más del 90% en muchos países) no recibe esos tratamientos. Uno de los principales obstáculos a una atención eficaz es la evaluación clínica errónea. En países de todo tipo de ingresos, las personas con depresión a menudo no son correctamente diagnosticadas, mientras

[5] El suicidio representa cerca de 1,5% de todas las defunciones en el mundo, por lo que se clasifica entre las 20 principales causas de muerte en el 2015 [10, p. 14]. El 79% de todos los suicidios se produce en países de ingresos bajos y medianos. La ingestión de plaguicidas, el ahorcamiento y las armas de fuego son algunos de los métodos más comunes de suicidio en todo el mundo [101].

que otras que en realidad no la padecen son a menudo diagnosticadas erróneamente y tratadas con antidepresivos. De hecho, un estudio [35] realizado por el Departamento de Psiquiatría de la Facultad de medicina de la Universidad Nacional de Colombia en 2014, con el fin de evaluar la efectividad de la detección[6] rutinaria para identificar pacientes con depresión, mostró que muchos casos positivos no son detectados. Se logró evidenciar, que los médicos de atención primaria (MAP) logran tasas de detección del trastorno depresivo entre 30% y 40% y por cada 100 personas atendidas hubo más falsos positivos que verdaderos positivos[7] o falsos negativos. Este estudio, ha determinado que algunos de los factores asociados al pobre reconocimiento de los cuadros depresivos son: pobre relación médico-paciente (baja confianza), baja destreza de comunicación del médico, dificultad del paciente para expresar su malestar, tiempo disponible de consulta, entre otros. Estos resultados han dejado a la vista el verdadero reto clínico que tiene la Atención Primaria de Salud (APS), ya que muchas personas que no padecen el trastorno están siendo sometidas a intervenciones innecesarias (ej. psicofármacos) y dando lugar a costos económicos innecesarios para los pacientes y el sistema de asistencia sanitaria. La tasa de detección de la depresión podría mejorar reduciendo su dependencia a la observación subjetiva de los médicos y proporcionando una evaluación mucho más objetiva y rápida.

En el campo de la Inteligencia Artificial (IA), los métodos de Aprendizaje Profundo o Deep Learning permiten extraer características de alto nivel y pueden usarse para crear sistemas que aprendan de las emociones y de los comportamientos humanos estrechamente relacionados con la depresión. Varias investigaciones [31] [32] han demostrado que las señales del habla en los pacientes con depresión y en la gente sin esta condición tienen diferencias significativas; Otros investigadores han aplicado métodos [33] [34] de

[6] El DSM-5 [102] y CIE-10 [103] incluyen criterios para hacer el diagnóstico de un trastorno depresivo.

[7] Falso positivo (diagnóstico positivo enfermedad ausente), Verdadero positivo (diagnóstico positivo enfermedad presente). [104]

Deep Learning[8] y [29] de Machine Learning[9] para el análisis del audio para identificar el grado de depresión de pacientes con depresión.

En el caso específico del modelo DepressionDetect, el cual usa un conjunto de datos de la Base de Datos DAIC-WOZ, compilada por el Instituto de Tecnologías Creativas de la USC y lanzada como parte AVEC 2017[10], este conjunto de datos consta de 189 sesiones, con un promedio de 16 minutos, entre un participante y una entrevistadora virtual llamada Ellie, controlada por un entrevistador humano en otra sala a través de un enfoque de "Mago de Oz". Antes de la entrevista, cada participante completó un cuestionario psiquiátrico (PHQ-8), del cual se deriva una clasificación de "verdad" si el entrevistado está depresivo, no depresivo.

El modelo se enfoca en las características prosódicas[11] para categorizar las características del habla, segmentando así el habla de la persona desde el silencio, el ruido y otros sonidos exteriores, toda esta información recopilada se usó para entrenar una red neuronal convolucional (CNN), para obtener los resultados sobre la detección de la depresión.

2.1.2. Enunciado del Problema:

Sobre la base de las consideraciones anteriores surge la siguiente pregunta:

¿El modelo DepressionDetect basado en redes Neuronales Convolucionales (CNN's) permite realizar un diagnóstico automático de la depresión mediante la voz?

[8] Deep Learning es un conjunto de algoritmos de aprendizaje automático (machine learning). [33]

[9] Machine Learning o aprendizaje automático, es un campo de las ciencias de la computación que abarca el estudio y la construcción de algoritmos capaces de aprender y hacer predicciones. Estas predicciones se pueden tomar como una clasificación de los datos de entrada a partir del reconocimiento de patrones existentes en los mismos [33]

[10] (AVEC 2017) sus siglas representan Desafío y Taller de Audio / Visual Emocional 2017 [4]

[11] Las características prosódicas pueden caracterizarse generalmente por un oyente e incluyen la longitud y el ritmo de las oraciones, la entonación y la frecuencia fundamental [4]

3. JUSTIFICACIÓN

Dado que la depresión es una condición común, incapacitante y que acorta la esperanza de vida de la persona que la padece [28, p. 2], es considerada una amenaza grave para la salud mental humana, donde esta puede afectar seriamente la vida normal de las personas, donde las personas con depresión pueden sentirse tristes, indefensas, vacías, ansiosas, anoréxicas, irritadas o molestas, y la depresión severa puede incluso conducir al suicidio [29, p. 2705], debe decirse que la motivación principal radica en la necesidad que se ha evidenciado en mejorar la detección y el diagnóstico de este trastorno mental de una manera temprana.

Se puede señalar a raíz de este problema que el presente proyecto se justifica en cómo pueden ser mejoradas la detección y el diagnóstico de la depresión en términos de detección automática mediante redes neuronales convolucionales. Este proyecto se realiza por que se vio una necesidad real en el diagnóstico ya que los primeros signos de depresión son difíciles de detectar y cuantificar, porque estos están limitados en gran medida por la observación subjetiva de los médicos y la falta de un diagnóstico de seguimiento a largo plazo, ya que a medida que aumenta el número de pacientes con depresión, esto pone una carga extra a los médicos para diagnosticar con precisión el grado de depresión [29, p. 2705].

Por otro lado, esta propuesta se hace para reducir la carga de los médicos para diagnosticar una gran cantidad de síntomas depresivos, respaldar los diagnósticos de los profesionales médicos, al igual que promover la remisión, prevenir la recaída y reducir la carga emocional de la enfermedad en los pacientes [4, p. 1].

Así, este proyecto contribuye desde el punto de vista social al campo de la salud al mejorar las condiciones de diagnóstico del paciente proporcionando una evaluación objetiva y un diagnóstico rápido a través de redes neuronales convolucionales, al igual que permite al autor obtener conocimientos significativos al ubicarse en el estado del arte con el proyecto en cuanto a la detección de la depresión se refiere, además brinda elementos fundamentales a la línea de investigación del grupo de investigación Nyquist[12].

[12] Nyquist: Grupo de investigación perteneciente al programa de Ingeniería de Sistemas y Computación adscrito a la Facultad de Ingenierías de la Universidad Tecnológica de Pereira.

4. OBJETIVO GENERAL Y ESPECÍFICOS

4.1. Objetivo General

Validar el Modelo DepressionDetect para el diagnóstico automático de la depresión mediante la voz basado en redes convolucionales profundas por medio de un despliegue en condiciones controladas.

4.2. Objetivos Específicos

- Hacer un estudio sobre la base de datos DAIC-WOZ, para identificar las características que permiten establecer el grado de depresión de una persona
- Realizar el procesamiento de los datos con base en la metodología utilizada en el modelo DepressionDetect
- Establecer a partir de la Base de Datos DAIC-WOZ el conjunto de datos de entrenamiento y el conjunto de datos de prueba que garantice una información sin sesgo
- Implementar la arquitectura del modelo DepressionDetect
- Entrenar el modelo implementado con el conjunto de datos de entrenamiento
- Comprobar el modelo implementado con el conjunto de datos de prueba disponibles en la base de datos.

5. MARCO DE REFERENCIA

5.1. Marco Teórico

En el marco del presente proyecto, es relevante e importante reconocer los conceptos básicos en los cuales está fundamentado el proyecto de manera general, al igual que saber más a nivel teórico sobre la depresión, por esta razón se plantea un paso por dichos conceptos, y se expresan a continuación.

5.1.1. Depresión, Una Vista General

La Depresión en sus inicios era conocida como melancolía[13], el origen del término aparece en diferentes textos o escritos de la antigüedad, el termino fue acuñado en 1725, cuando el británico Sir Richard Blackmore rebautiza el cuadro con el termino actual de depresión [36], alcanzando una denotación de enfermedad mental o un estado patológico, al ser este acompañado de otros síntomas, siendo calificado como tal una enfermedad que al ser un estado conflictivo suficientemente grave y duradero se pensó como en algo que tenía una entidad clínica.

Con el nacimiento de la psiquiatría moderna, una rama especializada de la medicina acuñada por Philippe Pinel, médico francés dedicado al estudio y tratamiento de las enfermedades mentales, y gracias a la biopsiquiatría y el despegue de la farmacología, la depresión se convierte en una enfermedad más susceptible de diagnóstico, tratamiento y de explicación bioquímica.

Su alta prevalencia y su relación con la esfera emocional la han convertido, a lo largo de la historia, en una condición común, incapacitante y que acorta la esperanza de vida de la persona que la padece, conllevando una amplia gama de problemas de salud mental caracterizados por la ausencia de afectividad positiva, es decir, una pérdida de interés o de la capacidad de disfrutar con las actividades que normalmente eran placenteras, el individuo también presenta un bajo estado de ánimo y una serie de síntomas emocionales, como pueden ser los sentimientos de culpa, de inutilidad, falta de ilusión y así como

[13] Según la Real Academia Española melancolía hace alusión a una tristeza vaga, profunda, sosegada y permanente, nacida de causas físicas o morales, que hace que quien la padece no encuentre gusto ni diversión en nada.

la baja autoestima con pérdida de confianza en sí mismos [37, p. 12].

Al estar la depresión constituida como una enfermedad, y saber que ha sido una condición que ha estado ligada a la humanidad por más de dos mil años, y que a lo largo del tiempo se han generado alrededor de ella cantidades de intentos por comprenderla desde su naturaleza y etiología[14], para generar así estrategias de abordaje desde la parte de diagnóstico y tratamiento. Se ha convertido y sigue estando vigente como unas de las entidades clínicas más desafiantes y desconcertantes para los profesionales de la Salud Mental.

5.1.1.1. Clasificación:

La depresión se puede clasificar según el número de síntomas, la intensidad de dichos síntomas y por los episodios depresivos que presenta una persona obteniendo así una primera clasificación que puede estar entre depresión leve, moderada o grave [38, p. 17]. Dependiendo el nivel de dicha depresión la persona puede pasar de presentar algunas dificultades al realizar algunas actividades laborales y sociales a ser incapaz de mantener dichas actividades, teniendo así que:

- Depresión Leve: Una persona con un episodio depresivo leve presenta un número pequeño de síntomas de depresión, y tiene una leve dificultad para llevar a cabo su actividad laboral y social.

- Depresión Moderada: Una persona en la categoría de depresión moderada tiene dificultades importantes para realizar su trabajo usual, así como sus actividades escolares, domésticas o sociales, debido a los síntomas de depresión.

- Depresión Grave: Durante un episodio depresivo grave la persona suele presentar una considerable angustia o agitación, hay un riesgo

[14] Según la Real Academia Española la etiología hace alusión al estudio de las causas de las enfermedades.

alto de suicidio en estos casos. El paciente suele tener una escasa capacidad para continuar con su actividad laboral, social o doméstica. Suele presentar síntomas físicos como dolor de cabeza, de espalda, entre otros que están asociados a la depresión.

Yendo a una clasificación más técnica tenemos:

- Depresión Endógena: Es la Depresión que se crea a dentro de nuestro cerebro, sin necesidad de que exista un factor externo que la produzca, está ligada y suele depender de cambios fisiológicos en el cerebro [39].

  Es un trastorno donde los pacientes son incapaces de sentir algo, no pueden sentir ira, empatía o felicidad por los sucesos que pasan en su entorno, de este tipo de depresión existen dos subtipos [37, p. 13] depresión unipolar[15] y trastorno afectivo bipolar[16].

- Depresión Exógena: Este tipo de Depresión depende de factores externos, se produce como consecuencia de acontecimientos externos, siendo las principales etiologías de este tipo:
  - Sucesos traumáticos como puede ser la muerte de un ser querido, una ruptura de pareja, perdida del trabajo entre otros sucesos.

  - Consecuencias de alguna enfermedad física.

- Depresión Orgánica: Este tipo de depresión es motivada por alguna causa orgánica como puede

[15] Depresión Unipolar: Durante los episodios depresivos típicos hay estado de ánimo deprimido, pérdida de interés y de la capacidad de disfrutar, y reducción de la energía que produce una disminución de la actividad, todo ello durante un mínimo de dos semanas. Muchas personas con depresión también padecen síntomas de ansiedad, alteraciones del sueño y del apetito, sentimientos de culpa y baja autoestima, dificultades de concentración e incluso síntomas sin explicación médica [37, p. 14].

[16] Trastorno Afectivo Bipolar: Este tipo de depresión consiste característicamente en episodios maníacos y depresivos separados por intervalos con un estado de ánimo normal. Los episodios maníacos cursan con estado de ánimo elevado o irritable, hiperactividad, autoestima excesiva y disminución de la necesidad de dormir [37, p. 14].

ser una patología, un fármaco, la falta de vitaminas o nutrientes en el organismo.

5.1.1.2. Síntomas:

Las personas presentan un grupo de alteraciones o síntomas emocionales, cognitivas, conductuales y psicológicas [40, pp. 417-449] relacionadas con la condición diagnóstica en este caso la depresión, que va acompañada además de estos síntomas de un bajo estado de ánimo. A continuación, se presenta una categorización de cada síntoma:

- Síntomas Emocionales: Entre estos se encuentran los sentimientos de culpa, de inutilidad, falta de ilusión, pérdida de confianza en sí mismos, animo bajo, irritabilidad.

- Síntomas Físicos y Conductuales: Pueden ser el llanto, el aislamiento social, la exacerbación de dolores preexistentes, fatiga, ansiedad marcada, disminución del sueño y del apetito, insomnio, perdida del deseo sexual.

- Síntomas Cognitivos: Pueden ser la pérdida de la concentración y reducción de la atención, pesimismo, pensamientos recurrentes negativos sobre uno mismo, enlentecimiento mental.

Junto a ellos las personas que padecen depresión suelen aludir al sentimiento de que todo les parece fútil o sin real importancia, acreditan que perdieron de forma irreversible la capacidad de sentir alegría o placer en la vida y todo les parece vacío o sin interés.

Ven el mundo sin color y sin alegría, ciertas personas depresivas también se presentan más apáticos que tristes y, a menudo, refieren sensación de falta de sentimientos (por ejemplo, no se inmutan ante el sufrimiento de un ser querido).

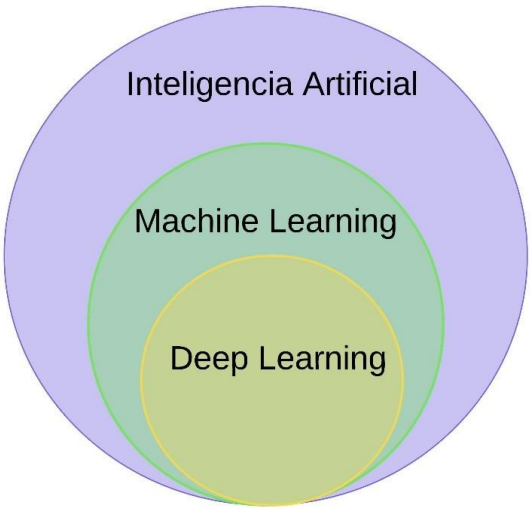
> Una persona depresiva, con frecuencia, se considera un peso para los familiares y amigos, invocando la muerte como forma de alivio o resolución de sus problemas, donde los pensamientos de suicidio varían desde el remoto deseo de estar simplemente muerto, hasta planos minuciosos y perfectamente armados de matarse [41, p. 69].

Ahora bien, después de haber dado una vista general por el concepto de depresión, procedemos a adentrarnos en los conceptos asociados a la Inteligencia Artificial y lo que esta conlleva.

5.1.2. Inteligencia Artificial

Hoy en día la Inteligencia Artificial tiene gran acogida y éxito en casi todas las ramas del conocimiento, la cual nació en los años 1950s, cuando un grupo de pioneros de la computación comenzaron a preguntarse si se podía hacer que las computadoras pensaran, para hablar de ella se parte de una definición, está definición la propuso Marvin Minsky, uno de los pioneros de la IA que dice que “la Inteligencia Artificial es la ciencia de construir máquinas para que hagan cosas, que, sí las hicieran los humanos, requerirían inteligencia” [42].

Entonces es posible pensar en la inteligencia artificial como en aquella ciencia que incorpora conocimiento a los procesos o actividades para que estos tengan éxito. La inteligencia artificial se divide en varias ramas, englobando a los campos de machine learning y Deep learning, como se observa en la Gráfica 1.



Gráfica 1 - Inteligencia Artificial (Desglose)

Fuente: Autoría propia, Edward Camilo Villota Taramuel

5.1.3. Machine Learning

Machine Learning, también conocido como aprendizaje automático, es una rama de la inteligencia artificial que abarca el estudio y la construcción de algoritmos capaces de aprender y hacer predicciones. Estas predicciones se pueden tomar como una clasificación de los datos de entrada a partir del reconocimiento de patrones existentes en los mismos [33].

Machine Learning nace de la necesidad de resolver cómo construir programas de computadora que mejoran automáticamente adquiriendo experiencia. Es decir, es una disciplina que abarca todos aquellos sistemas que aprenden automáticamente, donde dichos sistemas aprenden a identificar patrones complejos en multitud de datos, sin la necesidad de programar el conocimiento que son capaces de adquirir.

Volviéndose idóneo para el análisis de grandes volúmenes de datos, extrayendo y reconociendo patrones y tendencias para comprender que es lo que pueden decir esos datos, donde el aprendizaje automático se vale de algoritmos [43] [44] [45] [46] [47] que pueden procesar Gigas o Terabytes de datos y de ello obtener información útil. Todo esto se ha logrado gracias al gran avance de la mano de la tecnología y por el aumento de la

capacidad computacional de los computadores, permitiendo tratar problemas que anteriormente eran impensables.

El objetivo de los algoritmos de Machine Learning es construir un modelo que capture el conocimiento aprendido sobre los datos de entrada y que gracias a este modelo posteriormente se infiera conocimiento sobre nuevos datos, donde se intenta utilizar menos recursos para entrenar grandes volúmenes de datos e ir aprendiendo por sí mismos [48]. Machine Learning se subdivide en dos categorías, aprendizaje supervisado y aprendizaje no supervisado, cuya diferencia entre ambas es la forma en que aprende el modelo.

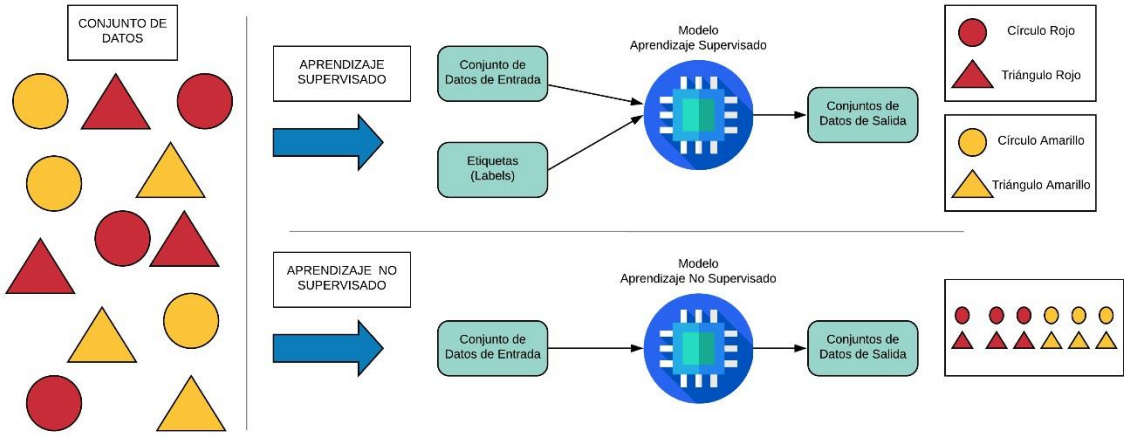
5.1.3.1. Clasificación en Machine Learning

- Aprendizaje Supervisado:

  En el Aprendizaje Supervisado los datos para el entrenamiento incluyen la solución deseada, llamada etiquetas o labels [49]**.** El modelo aprende a base de unos datos previamente etiquetados proporcionados por una persona, vale decir que en el modelo de aprendizaje supervisado una persona es la encargada de proporcionar las entradas, así como las salidas que debe dar el modelo para esos mismos datos, donde el modelo toma como entrada unos datos y devuelve un resultado o predicción sobre esos datos adaptándose para dar una salida esperada de acuerdo con los datos de entrada. Para la realización de este proyecto el enfoque se hace en el aprendizaje supervisado.

- Aprendizaje No Supervisado:

  En el aprendizaje no supervisado el modelo no parte de etiquetas o labels que tienen la solución deseada, el modelo intenta clasificar o descifrar la información por sí solo, considerando exclusivamente solo los datos de entrada [50].



Gráfica 2 - Aprendizaje Supervisado y Aprendizaje No Supervisado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

En la Gráfica 2, se describe de manera visual lo dicho anteriormente acerca del aprendizaje supervisado y el aprendizaje no supervisado.

5.1.4. Deep Learning

Deep Learning o aprendizaje profundo es un subcampo dentro de Machine Learning, que utiliza distintas estructuras de redes neuronales para lograr el aprendizaje por medio de sucesivas capas de representaciones cada vez más significativas de los datos [51]. En otros términos, es un conjunto de algoritmos de aprendizaje automático [33], esto consiste en llevar a cabo procesos de machine learning usando una red neuronal artificial que se compone de un número de niveles jerárquicos. En el nivel inicial de la jerarquía la red aprende algo simple y luego envía esa información al siguiente nivel. El siguiente nivel toma esta información sencilla, la combina, compone una información un poco más compleja, y se lo pasa al tercer nivel, y así sucesivamente [52, p. 12].

Deep o profundo en referencia a Deep Learning hace alusión a la cantidad de capas de representación que se utilizan en el modelo, la cantidad exacta de capas de representación varía dependiendo del tipo de modelo a implementar y los requerimientos de este, donde las capas de representación

aprenden automáticamente a medida que el modelo es entrenado con los datos.

Ahora bien, una manera de implementar Deep Learning es utilizar redes neuronales artificiales, donde también aparece el concepto de propagación hacia atrás, siendo estos conceptos tratados a continuación:

5.1.4.1. Redes Neuronales Artificiales [53, p. 9]

Una Red Neuronal está basada en el funcionamiento de las neuronas del ser humano y cómo ellas se comunican mediante estímulos eléctricos entre sí, siendo una red neuronal una herramienta matemática que modela, de forma muy simplificada, el funcionamiento de las neuronas en el cerebro.

Estos algoritmos o herramientas datan de los años 40 y 50, aunque en ese entonces no gozaron de una gran popularidad hasta estos días debido a la gran cantidad de recursos computacionales que requieren y de la enorme cantidad de datos de la que dependen para su correcto funcionamiento. Actualmente se les considera como una de las mejores técnicas para Deep Learning [50, p. 11].

- Estructura:

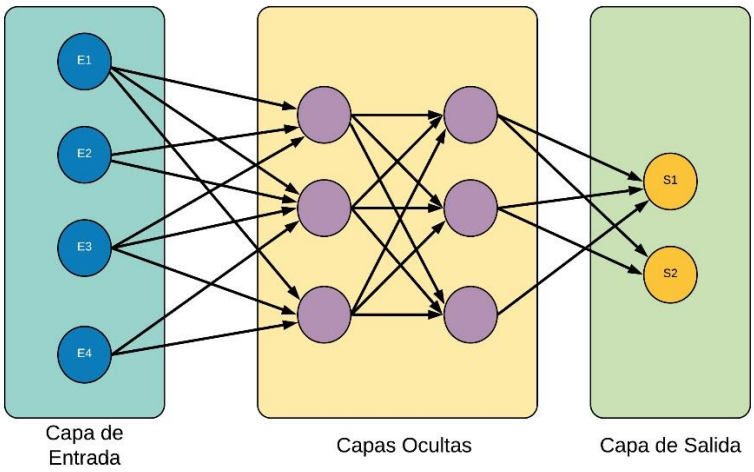
  En la Gráfica 3 es posible observar la arquitectura de una Red Neuronal, donde las entradas y salidas se conectan a otras neuronas que se llaman nodos, los nodos se encuentran fuertemente conectados entre sí, organizándose por capas teniendo así:

  - Capa de Entrada: Son aquellos nodos que reciben la información del exterior, reciben cada uno de los números de una lista entrante.

  - Capa de Salida: Son aquellos nodos que trasmiten la información al exterior, una vez que la red realiza su operación matemática,

transmite el resultado, también como una lista de números.

- Capa Oculta: Estos nodos no tienen contacto con el exterior y solamente intercambian información con otros nodos de la red, estos nodos adquieren un valor de los nodos de la capa de entrada que se van modificando en un proceso llamado aprendizaje, para de esa forma saber qué entrada es más importante o relevante, estos nodos contienen los cálculos intermedios de la red.



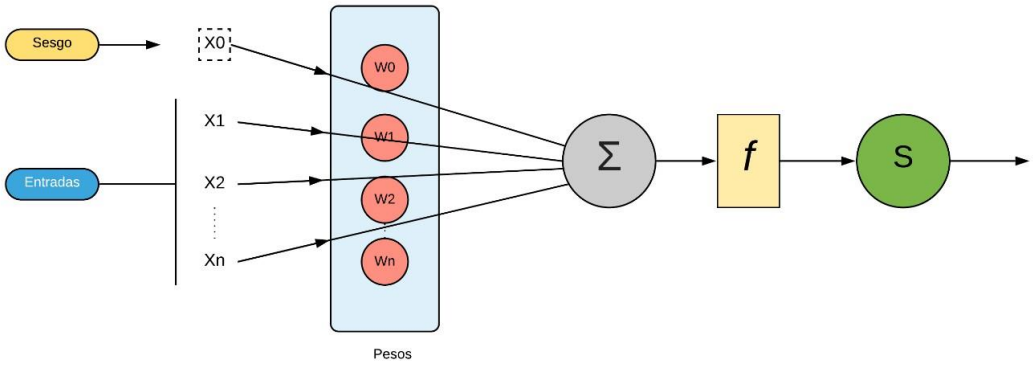
Gráfica 3 - Arquitectura Red Neuronal Artificial

Fuente: Autoría propia, Edward Camilo Villota Taramuel

- Funcionamiento:

  Cada neurona artificial o nodo, es una unidad de procesamiento que recibe una serie de señales de entrada que multiplica por un peso determinado llamado pesos sinápticos (W), el nodo calcula la suma del producto de cada entrada por su peso correspondiente y se aplica al valor resultante una función de activación que produce un valor de salida u otro, dependiendo de si la suma de señales y pesos supera un umbral determinado.

La estructura descrita anteriormente se puede observar de manera visual en la Gráfica 4:



Gráfica 4 - Esquema de una Red Neuronal (Pesos, Función de Activación)

Fuente: Autoría propia, Edward Camilo Villota Taramuel

En donde sus componentes son:

- X_1, X_2, …, X_n: Los datos de entrada del nodo o neurona, los cuales también pueden ser el producto de la salida de otro nodo de la red.

- X_0: Unidad de sesgo, es un valor constante que se le suma a la entrada de la función de activación del nodo, por lo general lleva el valor de 1.
  Este valor permite cambiar la función de activación de izquierda a derecha, otorgando más

flexibilidad en el momento de aprender del nodo o neurona.

- W_0, W_1, W_2, …, W_n: Son los pesos relativos o pesos sinápticos de cada entrada. También a la unidad de sesgo le corresponde un peso.

- f: Es la función de activación de la neurona, esta función es la que les otorga la flexibilidad a las redes neuronales.

- S: Es la salida de la neurona, que se calcula de la siguiente manera:

$$S = f(\sum_{i=0}^{n} w_i \, . \, x_i)$$

La red neuronal se organiza en capas de neuronas donde cada capa procesa la información de la anterior, el nivel de complejidad de los problemas que puede resolver la red se da por el número de capas y el tipo de función de activación.

5.1.4.2. Backpropagation (Propagación hacia Atrás) [54]

Es un algoritmo de entrenamiento el cual consiste en propagar el error de la capa de salida hacia las capas ocultas, lo cual da alusión a su traducción al español “retropropagación”, teniendo así un algoritmo que funciona mediante la determinación de la pérdida o error en la salida.

Entonces el término Backpropagation se refiere a la forma en que el error calculado en el lado de la capa de salida se propaga hacia atrás desde la capa de salida, a la capa oculta y finalmente a la capa de entrada. Cada una de las iteraciones en la retropropagación constituye dos barridos:

- Una activación hacia adelante para producir una solución

- Una propagación hacia atrás con el error calculado para modificar los pesos, de esta forma los pesos se van actualizando para minimizar el error resultante de cada nodo.

Los barridos hacia adelante y hacia atrás se realizan hasta que la solución esté de acuerdo con el valor deseado o esperado dentro de una tolerancia preespecificada. Así, este algoritmo permite a las redes neuronales aprender.

5.2. Marco Conceptual

En el presente marco conceptual del proyecto, es importante reconocer que los conceptos tratados a continuación posteriormente derivaran a reconocer a nivel conceptual la arquitectura a implementar en el proyecto.

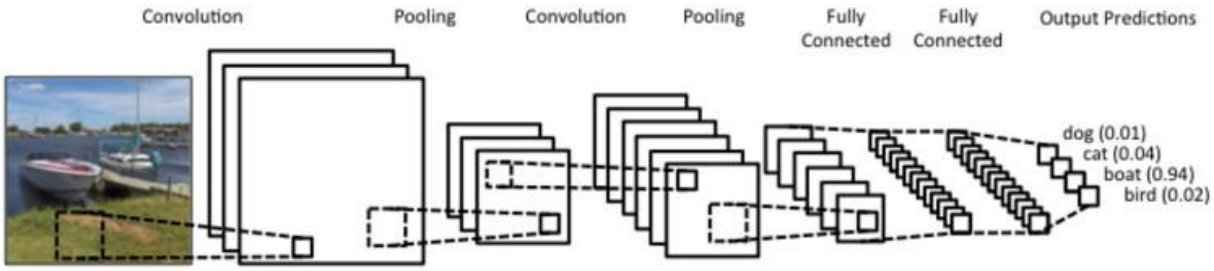
5.2.1. Redes Neuronales Convolucionales (CNN's)

Las CNN son un tipo particular de red neuronal inspiradas en el funcionamiento de la corteza visual del cerebro, son el tipo de redes más usadas en Deep Learning para tareas de reconocimiento de imagen por su gran rendimiento detectando objetos [55] [56], reconociendo patrones en imágenes [57], entre otras cosas [58] [59], demostrando ser así una herramienta poderosa en el reconocimiento de imágenes, el análisis de video y el procesamiento de lenguaje natural [4, p. 2].

Las redes neuronales convolucionales se asemejan bastante a las redes neuronales tradicionales o normales, ya que una CNN se compone de nodos o neuronas con pesos y sesgos que aprenden, usando sus entradas para realizar un producto escalar y luego aplicar una función de activación. La diferencia principal entre CNN y una red neuronal artificial es que en las CNN suponemos o tomamos como un hecho que las entradas a esta red son imágenes representadas en forma de matriz tridimensional, obteniendo con esto una ganancia de eficiencia y reducción de la cantidad de parámetros en la red.

Las redes CNN son redes feedforward, en las que la información fluye en un solo sentido, de las entradas a las salidas. La arquitectura de las redes neuronales convolucionales tiene muchas variantes; pero en general, consisten en capas convolucionales y pooling (submuestreo), las cuales se agrupan en módulos. Seguidos de una o más capas totalmente conectadas (fully connected), como en redes neuronales feedforward comunes. Los módulos se apilan uno sobre otro para formar un modelo profundo (Deep) [60].

5.2.1.1. Arquitectura General de una CNN



Gráfica 5 - Arquitectura general de la CNN [4, p. 2]

Las CNN se construyen utilizando cuatro tipos de capas principales: capa de entrada, capa convolucional, pooling y capa totalmente conectada (fully connected)

- Capa de Entrada: Son aquellos nodos que reciben la información del exterior.

- Capa Convolucional: En esta capa se aplica lo que se conoce como convolución que recibe la imagen como entrada y aplica sobre ella un filtro (kernel) que se usa normalmente para extraer características necesarias para hacer algunas operaciones ya que dependiendo del caso no se necesita todo lo que trae la imagen.

- Capa de Reducción (Pooling): Esta capa se ubica siempre después de una capa convolucional ya que esta capa tiene como fin, eliminar elementos innecesarios u obtener los elementos más representativos disminuyendo en este proceso la dimensión de la matriz tridimensional, disminuyendo su volumen de entrada para la siguiente capa.

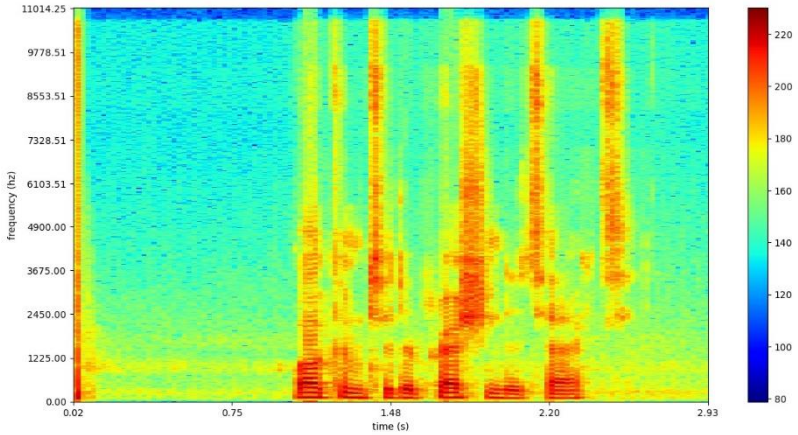
- Capa clasificadora totalmente conectada: Se encarga de calcular las puntuaciones obtenidas por la imagen de entrada para cada una de las clases o categorías definidas en el problema.

5.2.2. Espectrograma

Es una representación visual del sonido, que muestra la amplitud de los componentes de frecuencia de una señal a lo largo del tiempo, siendo los espectrogramas una representación gráfica del espectro de frecuencias de la emisión sonora. El espectrograma puede revelar rasgos, como altas frecuencias o modulaciones de amplitud, que no pueden apreciarse incluso aunque estén dentro de los límites de frecuencia del oído humano.

Los espectrogramas mantienen un alto nivel de detalle (incluido el ruido, que puede presentar desafíos para el aprendizaje de redes neuronales). La representación de un espectrograma se hace de la siguiente forma

Un espectrograma representa el tiempo sobre el eje horizontal, la frecuencia sobre el eje vertical y la amplitud de las señales mediante una escala de grises o de colores [61], como lo podemos observar en la Gráfica 6



Gráfica 6 - Espectrograma de un plosive, seguido de un segundo de silencio, y las palabras habladas, "Bienvenido a DepressionDetect " [4, p. 2].

5.2.3. Desequilibrio de Clase

El desequilibrio de clase se presenta cuando en el DataSet o conjunto de datos contamos con algunas muestras o clases desproporcionadas respecto a las otras, donde podría tener más o menor cantidad de información respecto a las demás,

esto provoca un desbalanceo en los datos que desea utilizar para el entrenamiento de la red.

Esto puede perjudicar a la red en el proceso de generalización de la información principalmente a las clases minoritarias. A manera de ejemplo si a una red neuronal se le dan 990 fotos de gatitos y sólo 10 de perros, no se puede pretender que logre diferenciar una clase de otra. Lo más probable es que la red se limite a responder siempre “tu foto es un gato” puesto que así tuvo un acierto del 99% en su fase de entrenamiento [48].

5.2.4. DAIC-WOZ Depression Database

Todas las grabaciones de audio y las métricas de depresión asociadas para la realización de este proyecto serán proporcionadas por la base de datos DAIC-WOZ, que fue compilada por el Instituto de Tecnologías Creativas de la USC y lanzada como parte del Desafío y Taller Audio / Visual Emocional 2017 (AVEC 2017). El conjunto de datos consta de 189 sesiones, con un promedio de 16 minutos, entre un participante y un entrevistador virtual llamado Ellie, controlado por un entrevistador humano en otra sala a través de un enfoque de "Mago de Oz". Antes de la entrevista, cada participante completó un cuestionario psiquiátrico (PHQ-8), del cual se deriva una clasificación binaria de "verdad" (depresivo, no depresivo) [4, p. 1].

5.2.5. Características Acústicas del Habla

5.2.5.1. Características Prosódicas

Las características prosódicas se refieren o hacen alusión a aquellos elementos que tienen que ver con rasgos de sonido, tonos y acentos. Estas características se manifiestan en las palabras para analizar su acentuación y la entonación general en la oración o frase. Estos elementos son importantes tanto para la organización del discurso como para la expresión de emociones por la combinación de los elementos: entonación, acentuación, ritmo y pausas [62, p. 34].

Para el desarrollo de este proyecto se van a tener en cuenta las características prosódicas como lo son el tono, ritmo,

acentuación, calidad de voz, articulación, entonación, la longitud y el ritmo de las oraciones.

5.2.6. Keras

Keras es una API de redes neuronales de alto nivel, escrita en Python y capaz de correr sobre frameworks como TensorFlow, CNTK, o Theano, para el desarrollo de este proyecto se hará uso de Keras para facilitar el proceso de experimentación rápida, además que admite redes convolucionales y permite la creación de prototipos fácilmente y de manera rápida.

5.2.7. Theano

Theano es una biblioteca de Python que permite definir, optimizar y evaluar expresiones matemáticas que involucran matrices multidimensionales de manera eficiente.

5.2.8. TensorFlow

Es una plataforma de código abierto de extremo a extremo para el aprendizaje automático. Cuenta con un ecosistema integral y flexible de herramientas, bibliotecas y recursos de la comunidad que les permite a los investigadores impulsar un aprendizaje automático innovador [25].

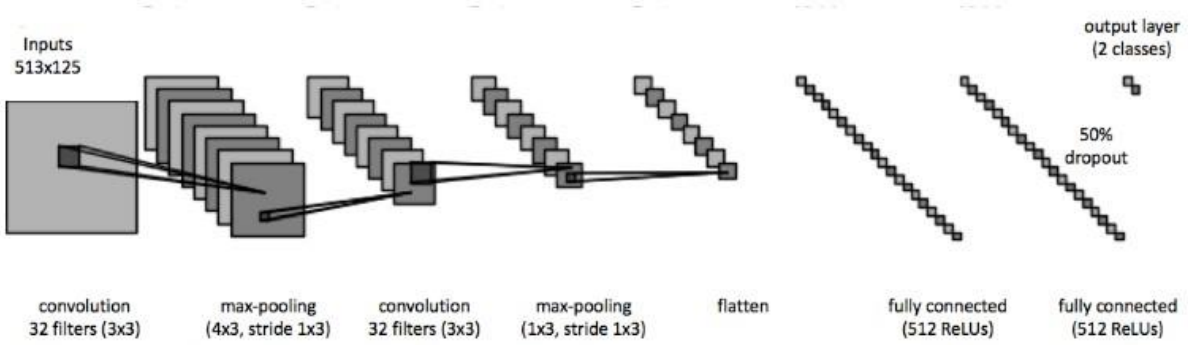
Su arquitectura flexible hace que pueda ser ejecutada en multitud de plataformas (GPUs, CPUs o TPUs) y en ordenadores, clústeres de servidores e incluso dispositivos móviles [63].

Una vez dado un recorrido por los conceptos claves del proyecto a nivel conceptual, se procede a desglosar la estructura y componentes que el proyecto va a tener:

5.2.9. Descripción del Modelo

El modelo de red neuronal convolucional (CNN) [4, p. 3] posee 6 capas que constan de 2 capas convolucionales con agrupación máxima y 2 capas totalmente conectadas.

Obteniendo la arquitectura de CNN como se observa en la Gráfica 7:



Gráfica 7 - DepressionDetect arquitectura CNN [4, p. 3]

La CNN comienza con una capa de entrada que se convierte en convolucionaria con filtros 32-3x3 para crear 32 mapas de características seguidos de una función de activación ReLU. A continuación, los mapas de características experimentan una reducción de dimensionalidad con una capa de agrupación máxima, que utilizara un filtro 4x3 con un paso de 1x3.

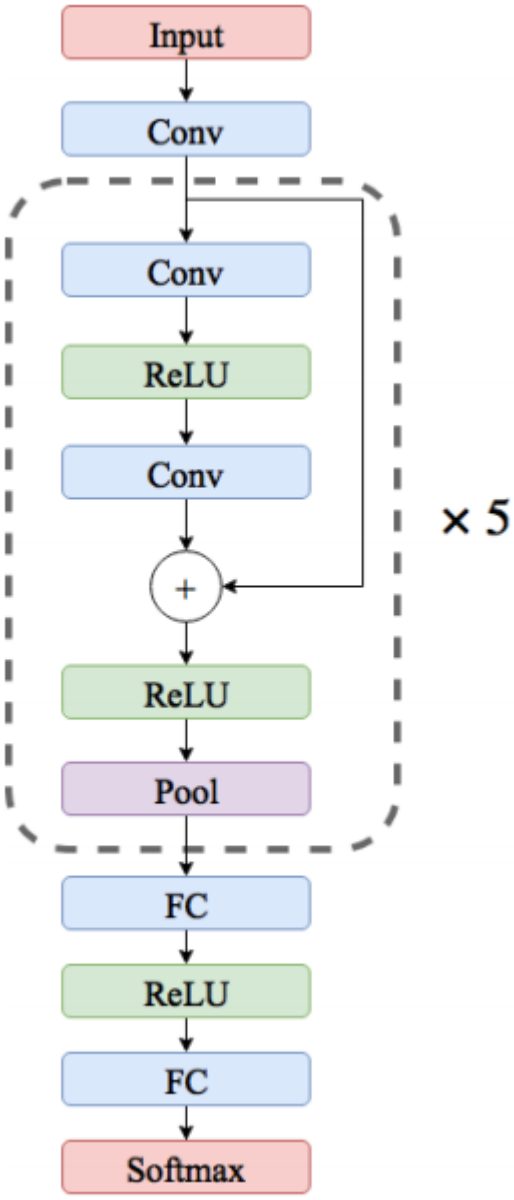
Una segunda capa convolucional similar se emplea con filtros de 32-3x3 seguidos de una capa de agrupación máxima con un filtro de 1x3 y zancada de 1x3.

Esta capa es seguida por dos capas densas. Después de la segunda capa densa, se usa una capa de deserción de 0.5.

Por último, se aplica una función softmax, que devuelve la probabilidad de estar en la clase depresiva o no depresiva. La suma de las probabilidades de cada clase es igual a 1.

Dicho modelo, nombrado anteriormente hace parte de la investigación principal, aunque se cambio por el modelo puesto a continuación, el cual varia diferentes parámetros como el numero de capas, el tamaño de los filtros, o las capas ocultas, teniendo un modelo más adecuado en relación a las entradas que se posee, permitiendo reducir el input de entrada cada vez que se realizan un par de operaciones convolución-agrupación.

Por ello, se empleará un modelo de red neuronal convolucional (CNN) que está compuesta por varias capas, implementando la extracción de funciones o características y posteriormente realizando la clasificación de la depresión [64, p. 3].



Gráfica 8 - Descripción de la Arquitectura del Modelo de la CNN [64, p. 3].

La CNN que se empleará comienza con una capa de entrada que se convierte en convolucional de 1-D con 32 filtros de tamaño 5, para crear 32 mapas de características.

Seguidos de cinco bloques residuales que contienen cada uno de ellos, una capa de convolución de 1-D con 32 filtros de tamaño 5 para crear 32 mapas de características seguidos de

una función de activación ReLU y una segunda capa convolucional similar con filtros de 32 de tamaño 5, seguidos de una función de activación ReLU. A continuación, los mapas de características experimentan una reducción de dimensionalidad con una capa de agrupación máxima (Max Pooling) de tamaño 5 con paso 2.

Por último, dos capas completamente conectadas (full-connected) o capas densas, acompañadas de una función de activación ReLU y una capa con una función softmax, que devuelve la probabilidad de que los datos ingresados esté en la clase depresiva o no depresiva.

5.3. Marco Legal

A continuación, se presentan un conjunto de leyes que regulan y protegen los derechos de las personas a cerca de poder tener acceso a la salud:

- Ley No. 1616 – 21 ene 2013 [65].

- Ley No. 1616 - 21 ene 2013, ARTICULO 1, OBJETIVO
  Garantizar el ejercicio pleno del Derecho a la Salud Mental a la población colombiana, priorizando a los niños, las niñas y adolescentes, mediante la promoción de la salud y la prevención del trastorno mental.

- Ley No. 1616 - 21 ene 2013, ARTICULO 3, SALUD MENTAL
  La Salud Mental es de interés y prioridad nacional para la República de Colombia, es un derecho fundamental, es tema prioritario de salud pública, es un bien de interés público y es componente esencial del bienestar general y el mejoramiento de la calidad de vida de colombianos y colombianas.

- Ley No. 1616 - 21 ene 2013, ARTICULO 6, Derechos de las Personas.

- Ley No. 1616 - 21 ene 2013, ARTICULO 7, De la promoción de la salud mental y prevención del trastorno mental.

- Articulo 49 - Constitución Política de Colombia
  La atención de la salud y el saneamiento ambiental son servicios públicos a cargo del Estado. Se garantiza a todas las personas el acceso a los servicios de promoción, protección y recuperación de la salud. Corresponde al Estado organizar, dirigir y reglamentar la prestación de servicios de salud a los habitantes y de saneamiento ambiental conforme a los principios de eficiencia, universalidad y solidaridad [66].

- Ley No. 1122 – 09 ene 2007, ARTICULO 1, OBJETIVO
  Realizar ajustes al Sistema General de Seguridad Social en Salud, teniendo como prioridad el mejoramiento en la prestación de los servicios a los usuarios. Con este fin se hacen reformas en los aspectos de dirección, universalización, financiación, equilibrio entre los actores del sistema, racionalización, y mejoramiento en la prestación de servicios de salud, fortalecimiento en los programas

de salud pública y de las funciones de inspección, vigilancia y control y la organización y funcionamiento de redes para la prestación de servicios de salud [67].

- Ley No. 100 – 23 dic 1993, ARTICULO 1: SISTEMA DE SEGURIDAD SOCIAL INTEGRAL.
  El sistema de seguridad social integral tiene por objeto garantizar los derechos irrenunciables de la persona y la comunidad para obtener la calidad de vida acorde con la dignidad humana, mediante la protección de las contingencias que la afecten [68].

- Acuerdo sobre Inteligencia Artificial ante la OCDE
  La IA debería beneficiar a las personas y al planeta impulsando el crecimiento inclusivo, el desarrollo sostenible y el bienestar [69].

6. ESTADO DEL ARTE

La depresión es un trastorno del estado de ánimo típico, que afecta a un número significativo de personas en todo el mundo a un ritmo creciente [70]. Se han aplicado diferentes métodos de aprendizaje automático para evaluar los estados de la depresión [71] y se están realizando investigaciones activas [72] para mejorar aún más la fiabilidad de esos sistemas en base a la necesidad de buscar un método efectivo para el diagnóstico de este padecimiento.

Hoy en día el avance de la tecnología y el aumento de la capacidad computacional de los ordenadores, ha permitido llegar a un mayor nivel de automatización y eficiencia en el proceso de validación y articulación de diferentes síntomas que puede presentar una persona con este tipo de enfermedad, viendo en campos como [29] Machine learning y [33] [34] Deep learning, campos prometedores para realizar el diagnóstico de la Depresión.

Al plantear una visión más profunda en relación con el diagnóstico de la depresión, se ha estado trabajando en dos grandes ejes, identificado así dos grandes temáticas a las cuales se ha ido orientado el diagnóstico automático de la Depresión.

Primero, los estudios entorno al reconocimiento de la depresión mediante imágenes, en 2017 un estudio publicado en la revista EPJ Data Science, donde Reece A. y Danforth C. emplean una metodología para analizar datos fotográficos de Instagram para detectar de forma predictiva la depresión [73], empleando aprendizaje automático, pudiendo detectar con éxito personas depresivas a partir de pistas en sus fotos de Instagram; otras investigaciones [74] [75] redireccionadas al apartado de imágenes, donde hacen un enfoque a los signos visuales desde la perspectiva del procesamiento de imágenes, donde buscan estar más cerca a obtener un método para el análisis automatizado de la depresión, que pueda ayudar a los médicos en el diagnóstico y monitoreo de la misma. Y por último, los estudios entorno al reconocimiento de la depresión mediante la voz, encontrándose en las técnicas de Inteligencia Artificial para hacer análisis de la depresión mediante la voz al igual que en las imágenes, un escenario prometedor para un diagnostico óptimo.

El habla como modalidad para realizar la detección de la depresión ha sido una de las principales modalidades exploradas, teniendo un alto nivel de atención el estudio de la evaluación automática, varias investigaciones [31] [32] han demostrado que las señales del habla en los pacientes con depresión y en la gente sin esta condición tienen diferencias significativas, teniendo así que los métodos de aprendizaje automático y aprendizaje profundo pueden usarse para aprender emociones y comportamientos expresivos directamente relacionados con la depresión.

Los modelos de redes neuronales artificiales recientes y los métodos de análisis automático de emociones para problemas relacionados con la depresión son extensos [76] [77] [78] teniendo así videncia de que ciertos parámetros vocales

pueden usarse aún más para discriminar objetivamente el habla depresiva y la que no lo es, como sostienen France, D. J *et al.* [31], que la producción del habla en pacientes con depresión es diferente a la de las personas normales.

Diferentes autores han optado por trabajar con las señales de audio como una herramienta para diagnosticar la depresión [79] [80] [81], ya que la idea de reconocer el estado afectivo de las personas para el diagnóstico y monitoreo de la salud es prometedora y lo seguirá siendo en un futuro para la atención médica. Esto permite el monitoreo de la salud a largo plazo, que es importante para el tratamiento y manejo no solo de la depresión sino de una amplia gama de enfermedades crónicas, trastornos neurológicos y problemas de salud mental [72, p. 2].

7. DISEÑO METODOLÓGICO

7.1. Hipótesis

El modelo DepressionDetect permite validar la posibilidad de detección automática de la depresión de una persona mediante la voz.

7.2. Tipo de investigación

Para el presente proyecto, la metodología propuesta es de tipo cuantitativo haciendo uso del método experimental ya que vamos a emplear una forma estructurada para la recopilación y análisis de la información.

7.3. Población

La población o universo para este estudio son los datos de la Base de Datos DAIC-WOZ, en la que los sujetos comparten ciertas características que permiten observar el grado de depresión de una persona, teniendo así que las fuentes de información es decir el conjunto de datos con la que se trabajará a lo largo del proyecto, serán obtenidos de dicha base de datos, compilada por el Instituto de Tecnologías Creativas de la USC y lanzada como parte AVEC 2017, este conjunto de datos consta de 189 sesiones, con un promedio de 16 minutos, entre un participante y una entrevistadora virtual llamada Ellie, controlada por un entrevistador humano en otra sala a través de un enfoque de "Mago de Oz", esta es una Base de Datos especializada para respaldar el diagnóstico de trastornos psicológicos como la ansiedad, la depresión y el estrés postraumático.

7.4. Muestra

Al momento de seleccionar la muestra y teniendo en cuenta los datos agrupados en la Base de Datos DAIC-WOZ, el número de sujetos no depresivos es aproximadamente cuatro veces mayor que el de los sujetos depresivos, lo que puede presentar un sesgo de clasificación como "no depresivos", además se puede producir otro tipo de sesgo al enfatizar en algunas características que son específicas de cada persona al tener las entrevistas un rango de duración que varía entre 7 a 33 minutos.

Por esa razón, la muestra tomada de los participantes que se encuentran en la base de datos se le aplica una serie de procedimientos como son: balanceo de la información de la base de datos, un submuestreo de las muestras de audio, seguida de un remuestreo de las mismas para tener igual cantidad de participantes depresivos y no depresivos, tomando

muestras en una proporción de 50/50 de cada clase (depresivo, no depresivo), para el conjunto de datos de entrenamiento y prueba.

7.5. Variables

Para la realización del cálculo de todas las variables encontradas en el diseño de instrumentos de realizan los siguientes cálculos

		Predicción	
		Positivos	**Negativos**
Observación	**Positivos**	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Gráfica 9 - Estructura Matriz de Confusión [82]

Partimos de una estructura de matriz de confusión[17] ver Gráfica 9, para calcular los siguientes valores:

- F1 score: Es necesaria cuando se desea buscar un equilibrio entre precision y recall (sensibilidad), pudiéndose calcular con la siguiente formula:

$$F1\ score = 2 * \frac{precision * recall}{precision + recall}$$

- Precision: Este apartado establece que del total de positivos cuantos son realmente positivos, viendo que tan bueno es el modelo para clasificar positivos, pudiéndose calcular con la siguiente formula:

$$precision = \frac{VP}{Total\ Clasificados\ Positivos\ (VP + FP)}$$

- Recall: Este apartado permite ver, cuándo la clase es positiva, que porcentaje se logró clasificar correctamente, donde se realiza una

[17] Matriz de Confusión: Es una herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado, una matriz nxn en la que las filas se nombran según las clases reales y las columnas, según las clases previstas por un modelo, sirviendo para mostrar de forma explícita cuándo una clase es confundida con otra [113].

división entre los verdaderos positivos y el total de positivos, obteniendo la siguiente formula:

$$recall = \frac{VP}{Total\ Positivos\ (VP + FN)}$$

- Accuracy: La exactitud se mide estableciendo cuántos se clasificaron como verdaderos positivos más cuántos se clasificaron como verdaderos negativos, dividiendo eso por el total teniendo la siguiente formula:

$$accuracy = \frac{VP + VN}{Total\ (VP + VN + FP + FN)}$$

7.6. Diseño de instrumentos para toma de información

Debido al tamaño de los datos no fue posible correr el modelo con todos estos por lo que se limitó el tamaño de muestra a 100.000. Sin embargo, como un elemento de verificación se corrió el modelo con muestras de este tamaño en ubicaciones diferentes de los datos, específicamente en los rangos 0 a 100.000, 1.000.000 a 1.100.000, 2.000.000 a 2.100.000, 3.000.000 a 3.100.000, 4.000.000 a 4.100.000, en base a los cuales se obtuvieron los resultados, lo que conlleva que para el diseño de instrumentos se va a analizar Tabla 1, Tabla 2, Tabla 3, Tabla 4, Tabla 5, Tabla 6, Tabla 7, Tabla 8, Tabla 9, Tabla 10:

Tabla 1 - Predicciones de conjuntos de prueba para el rango de 0 a 100.000 muestras de audio

Tabla 1: Predicciones de conjuntos de prueba para el rango de 0 a 100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	Verdaderos Positivos (VP)	Falsos Negativos (FN)
Pronosticada: No	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Para un conjunto de prueba con rango de 0 a 100.000 muestras de audio se dispone de una matriz de confusión (Confusion Matrix), la cual se encarga de evaluar la calidad de la predicción del modelo.

Donde tenemos:

- Verdaderos Positivos (VP): Es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.
- Verdaderos Negativos (VN): Es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.
- Falsos Negativos (FP): Es la cantidad de positivos que fueron clasificados incorrectamente como negativos.
- Falsos Positivos (FN): Es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

Tabla 2 - Predicciones de conjuntos de prueba para el rango de 0 a 100.000 muestras de audio

Tabla 2: Predicciones de conjuntos de prueba para el rango de 0 a 100.000 muestras de audio			
F1 score X_1	precision X_2	recall X_3	accuracy X_4

Para el conjunto de prueba con rango de 0 a 100.000 muestras de audio disponemos de lo siguiente donde X_1, X_2, X_3, X_4 corresponde a:

- X_1: Es necesaria cuando se desea buscar un equilibrio entre precision y recall (sensibilidad), teniendo así una medida de la precisión de la prueba.
- X_2: Este aparatado nos dice que del total de positivos cuantos son realmente positivos, viendo que tan bueno es el modelo para clasificar positivos.
- X_3: Este apartado nos ayuda a ver cuándo la clase es positiva, que porcentaje se logró clasificar correctamente.
- X_4: Es el porcentaje de los datos que se clasifican correctamente.

Tabla 3 - Predicciones de conjuntos de prueba para el rango de 1.000.000 a 1.100.000 muestras de audio

Tabla 3: Predicciones de conjuntos de prueba para el rango de 1.000.000 a 1.100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	Verdaderos Positivos (VP)	Falsos Negativos (FN)

Pronosticada: No	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Para un conjunto de prueba con rango de 1.000.000 a 1.100.000 muestras de audio disponemos de una matriz de confusión (Confusion Matrix), la cual se encarga de evaluar la calidad de la predicción del modelo.

Donde tenemos:

- Verdaderos Positivos (VP): Es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.
- Verdaderos Negativos (VN): Es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.
- Falsos Negativos (FP): Es la cantidad de positivos que fueron clasificados incorrectamente como negativos.
- Falsos Positivos (FN): Es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

Tabla 4 - Predicciones de conjuntos de prueba para el rango de 1.000.000 a 1.100.000 muestras de audio

Tabla 4: Predicciones de conjuntos de prueba para el rango de 1.000.000 a 1.100.000 muestras de audio			
F1 score X_1	precision X_2	recall X_3	accuracy X_4

Para el conjunto de prueba con rango de 1.000.000 a 1.100.000 muestras de audio disponemos de lo siguiente donde X_1, X_2, X_3, X_4 corresponde a:

- X_1: Es necesaria cuando se desea buscar un equilibrio entre precision y recall (sensibilidad), teniendo así una medida de la precisión de la prueba.
- X_2: Este aparatado nos dice que del total de positivos cuantos son realmente positivos, viendo que tan bueno es el modelo para clasificar positivos.
- X_3: Este apartado nos ayuda a ver cuándo la clase es positiva, que porcentaje se logró clasificar correctamente.
- X_4: Es el porcentaje de los datos que se clasifican correctamente.

Tabla 5 - Predicciones de conjuntos de prueba para el rango de 2.000.000 a 2.100.000 muestras de audio

Tabla 5: Predicciones de conjuntos de prueba para el rango de 2.000.000 a 2.100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	Verdaderos Positivos (VP)	Falsos Negativos (FN)
Pronosticada: No	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Para un conjunto de prueba con rango de 2.000.000 a 2.100.000 muestras de audio disponemos de una matriz de confusión (Confusion Matrix), la cual se encarga de evaluar la calidad de la predicción del modelo.

Donde tenemos:

- Verdaderos Positivos (VP): Es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.
- Verdaderos Negativos (VN): Es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.
- Falsos Negativos (FP): Es la cantidad de positivos que fueron clasificados incorrectamente como negativos.
- Falsos Positivos (FN): Es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

Tabla 6 - Predicciones de conjuntos de prueba para el rango de 2.000.000 a 2.100.000 muestras de audio

Tabla 6: Predicciones de conjuntos de prueba para el rango de 2.000.000 a 2.100.000 muestras de audio			
F1 score X_1	precision X_2	recall X_3	accuracy X_4

Para el conjunto de prueba con rango de 2.000.000 a 2.100.000 muestras de audio disponemos de lo siguiente donde X_1, X_2, X_3, X_4 corresponde a:

- X_1: Es necesaria cuando se desea buscar un equilibrio entre precision y recall (sensibilidad), teniendo así una medida de la precisión de la prueba.
- X_2: Este aparatado nos dice que del total de positivos cuantos son realmente positivos, viendo que tan bueno es el modelo para clasificar positivos.
- X_3: Este apartado nos ayuda a ver cuándo la clase es positiva, que porcentaje se logró clasificar correctamente.
- X_4: Es el porcentaje de los datos que se clasifican correctamente.

Tabla 7 - Predicciones de conjuntos de prueba para el rango de 3.000.000 a 3.100.000 muestras de audio

Tabla 7: Predicciones de conjuntos de prueba para el rango de 3.000.000 a 3.100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	Verdaderos Positivos (VP)	Falsos Negativos (FN)
Pronosticada: No	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Para un conjunto de prueba con rango de 3.000.000 a 3.100.000 muestras de audio disponemos de una matriz de confusión (Confusion Matrix), la cual se encarga de evaluar la calidad de la predicción del modelo.

Donde tenemos:

- Verdaderos Positivos (VP): Es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.
- Verdaderos Negativos (VN): Es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.
- Falsos Negativos (FP): Es la cantidad de positivos que fueron clasificados incorrectamente como negativos.
- Falsos Positivos (FN): Es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

Tabla 8 - Predicciones de conjuntos de prueba para el rango de 3.000.000 a 3.100.000 muestras de audio

Tabla 8: Predicciones de conjuntos de prueba para el rango de 3.000.000 a 3.100.000 muestras de audio			
F1 score X_1	precision X_2	recall X_3	accuracy X_4

Para el conjunto de prueba con rango de 3.000.000 a 3.100.000 muestras de audio disponemos de lo siguiente donde X_1, X_2, X_3, X_4 corresponde a:

- X_1: Es necesaria cuando se desea buscar un equilibrio entre precision y recall (sensibilidad), teniendo así una medida de la precisión de la prueba.
- X_2: Este aparatado nos dice que del total de positivos cuantos son realmente positivos, viendo que tan bueno es el modelo para clasificar positivos.
- X_3: Este apartado nos ayuda a ver cuándo la clase es positiva, que porcentaje se logró clasificar correctamente.
- X_4: Es el porcentaje de los datos que se clasifican correctamente.

Tabla 9 - Predicciones de conjuntos de prueba para el rango de 4.000.000 a 4.100.000 muestras de audio

Tabla 9: Predicciones de conjuntos de prueba para el rango de 4.000.000 a 4.100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	Verdaderos Positivos (VP)	Falsos Negativos (FN)
Pronosticada: No	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Para un conjunto de prueba con rango de 4.000.000 a 4.100.000 muestras de audio disponemos de una matriz de confusión (Confusion Matrix), la cual se encarga de evaluar la calidad de la predicción del modelo.

Donde tenemos:

- Verdaderos Positivos (VP): Es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.
- Verdaderos Negativos (VN): Es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.
- Falsos Negativos (FP): Es la cantidad de positivos que fueron clasificados incorrectamente como negativos.
- Falsos Positivos (FN): Es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

Tabla 10 - Predicciones de conjuntos de prueba para el rango de 4.000.000 a 4.100.000 muestras de audio

Tabla 10: Predicciones de conjuntos de prueba para el rango de 4.000.000 a 4.100.000 muestras de audio			
F1 score X_1	precision X_2	recall X_3	accuracy X_4

Para el conjunto de prueba con rango de 4.000.000 a 4.100.000 muestras de audio disponemos de lo siguiente donde X_1, X_2, X_3, X_4 corresponde a:

- X_1: Es necesaria cuando se desea buscar un equilibrio entre precision y recall (sensibilidad), teniendo así una medida de la precisión de la prueba.
- X_2: Este aparatado nos dice que del total de positivos cuantos son realmente positivos, viendo que tan bueno es el modelo para clasificar positivos.
- X_3: Este apartado nos ayuda a ver cuándo la clase es positiva, que porcentaje se logró clasificar correctamente.
- X_4: Es el porcentaje de los datos que se clasifican correctamente.

7.7. Descripción metodológica del proceso de desarrollo de cada uno de los objetivos específicos.

Para el proceso de desarrollo de cada uno de los objetivos específicos, se realiza una descripción de cómo se van a desarrollar estos, con todas las actividades o técnicas que requieren para llevarlos a cabo:

- Hacer un estudio sobre la base de datos DAIC-WOZ, para identificar las características que permiten establecer el grado de depresión de una persona
  - Solicitar la base de datos DAIC-WOZ.

- Firmar y enviar el formulario de acuerdo para la obtener la base de datos DAIC-WOZ.
- Descargar la base de datos DAIC-WOZ.
- Comprender el desglose de la estructura de la base de datos DAIC-WOZ y sus conjuntos de datos como son train, test, dev.
- Análisis de las Transcripciones de las entrevistas de los participantes de la base de datos DAIC-WOZ.
- Análisis de PHQ-8 y cómo este influye al detectar la depresión.
- Establecer las características prosódicas que se usaran como predictores prometedores de la depresión.
- Verificar la calidad de los archivos de audio.

- Realizar el procesamiento de los datos con base de la metodología utilizada en el modelo DepressionDetect
  - Análisis previo de los audios dispuestos por la Base de Datos.
  - Descargar e instalar Anaconda: Una plataforma Open Source utilizada para ciencia de datos y aprendizaje automático.
  - Segmentación del audio de los Participantes de la Base de Datos.
  - Balanceo de la Base de Datos sin pérdida de información.

- Establecer a partir de la Base de Datos DAIC-WOZ el conjunto de datos de entrenamiento y el conjunto de datos de prueba que garantice una información sin sesgo
  - Remuestreo de las muestras de audio de los participantes, tomando un número fijo de segmentos de cada uno de los participantes para garantizar que la CNN tenga la misma duración de entrevista para cada participante.
  - Asignación del conjunto de datos de entrenamiento y prueba en una proporción de 50/50 de cada clase (depresivo, no depresivo) para cada conjunto.

- Implementar la arquitectura del modelo DepressionDetect
  - Descargar e instalar TensorFlow
  - Descargar e instalar Keras.
  - Descargar e instalar Theano.
  - Implementar el modelo de red neuronal convolucional (CNN) utilizando Keras, TensorFlow y Theano.

- Entrenar el modelo implementado con el conjunto de datos de entrenamiento

- Cargar el conjunto de datos de entrenamiento dispuesto para la CNN
- Seleccionar el conjunto de datos de entrenamiento con los participantes de cada categoría de depresión (depresivo, no depresivo).
- Entrenar el modelo con los conjuntos de datos correspondientes de cada rango establecido durante 50 épocas.

- Comprobar el modelo implementado con el conjunto de datos de prueba disponibles en la base de datos.
  - Probar el modelo con los datos de prueba.
  - Identificar la precisión de las predicciones del conjunto de prueba.
  - Identificar la exactitud de las predicciones del conjunto de prueba.
  - Comparar los resultados obtenidos con los de anteriores investigaciones.
  - Concluir acerca de la capacidad del modelo DepressionDetect para el diagnóstico automático de la depresión mediante voz.

8. IMPLEMENTACIÓN

8.1. Hacer un estudio sobre la base de datos DAIC-WOZ, para identificar las características que permiten establecer el grado de depresión de una persona

8.1.1. Levantamiento de Información

8.1.1.1. Solicitud de la Base de Datos DAIC-WOZ

Se solicito la base de datos DAIC-WOZ, mediante la firma y envió de un formulario de acuerdo, a la siguiente dirección de correo electrónico boberg@ict.usc.edu, para obtener dicha base de datos.

Debido a restricciones de consentimiento, solo se permitió acceder a los datos para fines académicos e investigativos sin fines de lucro, siendo esto idóneo para el proyecto.

El envío del formulario de acuerdo, se diligencio mediante la dirección de correo institucional al solicitar la descarga de datos.

8.1.1.2. Descarga de la Base de datos DAIC-WOZ

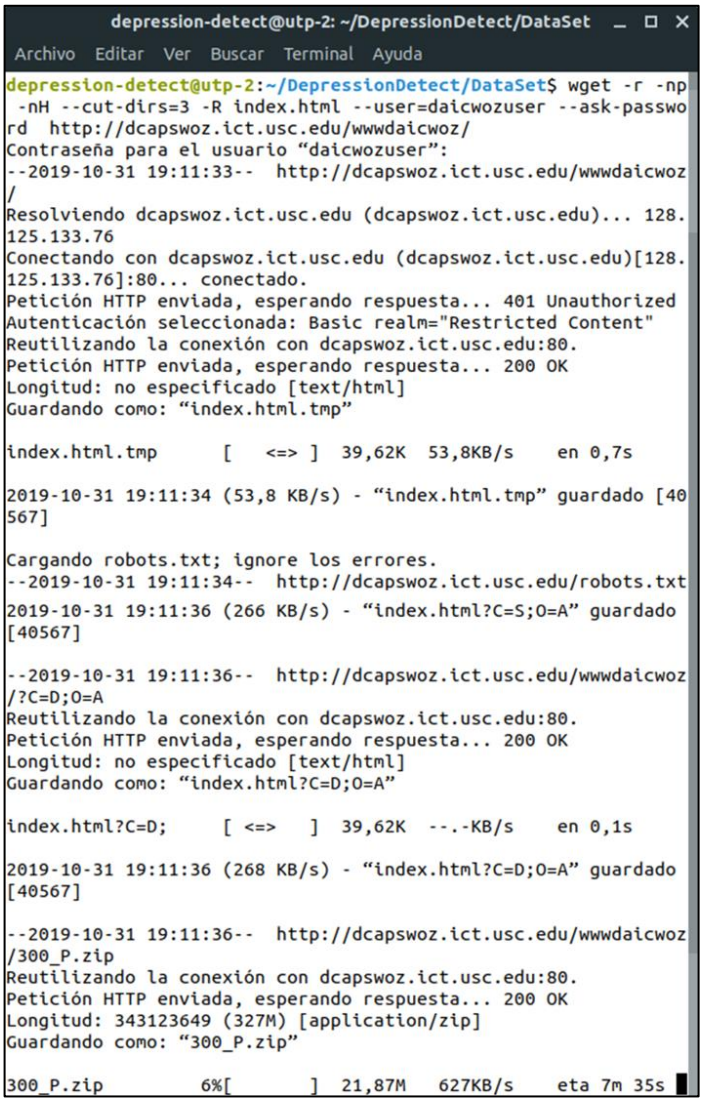
Al tener la aceptación por parte del Instituto de Tecnologías Creativas de la USC y poder acceder a los datos lanzados como parte del Desafío y Taller Audio / Visual Emocional 2017 (AVEC 2017).

Los cuales brindaron un nombre de usuario y contraseña respectiva para descargar los datos correspondientes, los datos son un paquete que contiene 189 archivos .zip donde cada uno contiene una carpeta de sesiones que va numerada desde 300- 492, además de la bibliografía de la base de datos contenida en documents.zip y archivos .csv con el listado de los conjuntos de datos para train, dev y test.

- Para descargar los archivos .zip de la base de datos desde nuestra terminal se introdujo el siguiente comando:

```
wget -r -np -nH --cut-dirs = 3 -R index.html --user =
daicwozuser --ask-password
http://dcapswoz.ict.usc.edu/wwwdaicwoz/
```

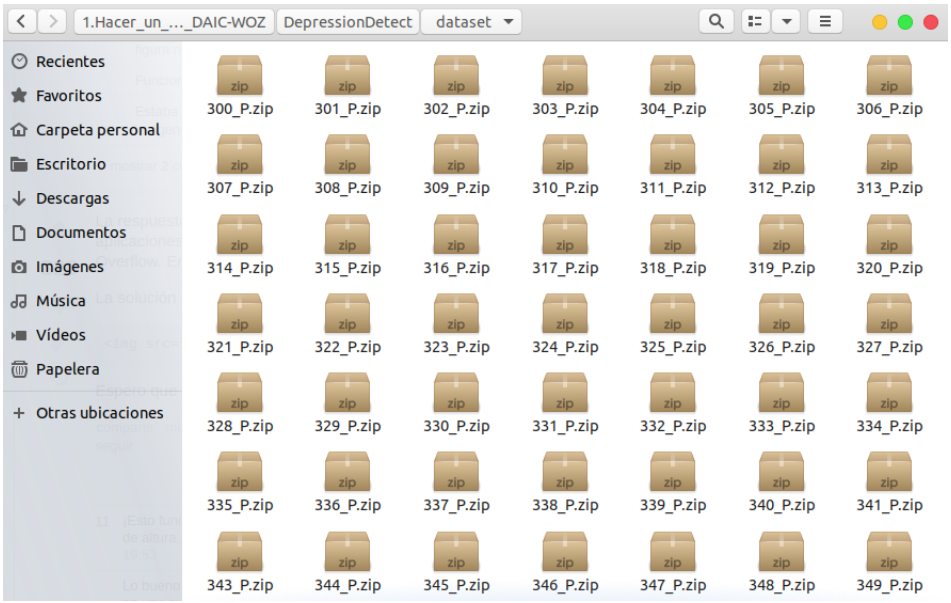
Donde se debe introducir la contraseña para el usuario “daicwozuser” y posteriormente a eso se procede a descargar cada uno de los datos de DAIC-WOZ, en la Gráfica 10 es posible observar dicho proceso.



Gráfica 10 - Descarga de la Base de Datos DAIC-WOZ

Fuente: Autoría propia, Edward Camilo Villota Taramuel

- En el entorno de trabajo, los archivos .zip quedaron de la siguiente forma como lo podemos observar en la Gráfica 11.

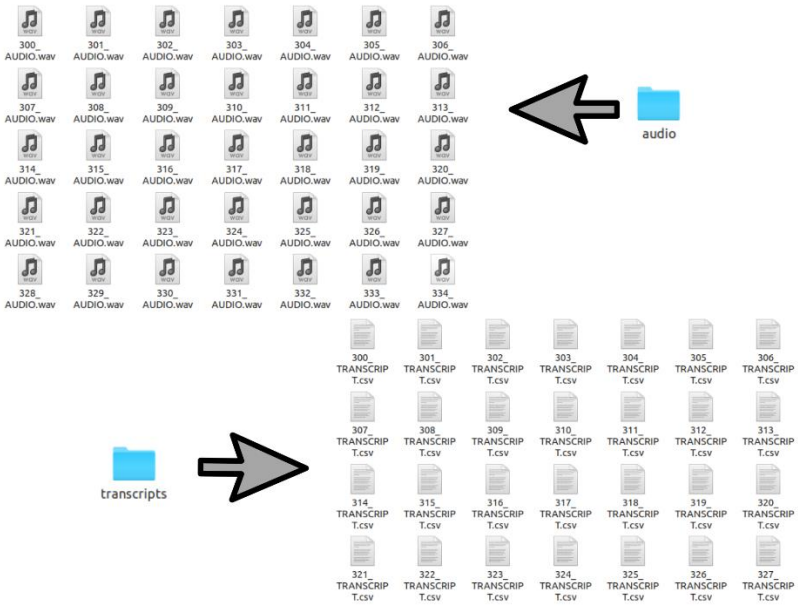


Gráfica 11 - Conjunto de Datos DAIC-WOZ

Fuente: Autoría propia, Edward Camilo Villota Taramuel

- Para descomprimir los archivos .zip con las entrevistas contenidas en la base de datos DAIC-WOZ, se empleó un código realizado en Python en un notebook de Jupyter Notebook[18] llamado Descomprimir_Dataset_Zip.ipynb, para cumplir dicha tarea, dejando la información almacenada en dos carpetas, una de ellas llamada `audio`, la cual contiene los archivos .wav con las entrevistas de los participantes y `transcripts`, la cual contiene los archivos .csv con las transcripciones de las entrevistas de los participantes, observadas en la Gráfica 12:

[18] Jupyter Notebook: Es una aplicación web de código abierto que permite crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo, Jupyter admite más de 40 lenguajes de programación, incluidos Python, R, Julia y Scala [108].



Gráfica 12 - Audios y Transcripts - Base de Datos DAIC-WOZ

Fuente: Autoría propia, Edward Camilo Villota Taramuel

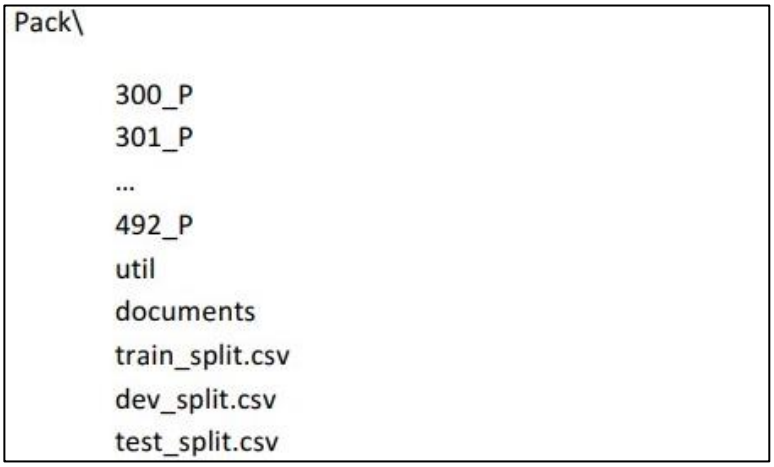
8.1.2. Desglose de la estructura de la base de datos DAIC-WOZ

8.1.2.1. Estructura de la Base de datos DAIC-WOZ

La Base de datos está constituida por un conjunto de datos que consta de 189 sesiones, con un promedio de 16 minutos, dichas sesiones o entrevistas están recopiladas como parte de un esfuerzo mayor para crear un agente informático que entreviste a las personas e identifique indicadores verbales y no verbales de enfermedad mental.

Los datos recopilados incluyen grabaciones de audio y video y amplias respuestas al cuestionario PHQ-8.

Cada sesión incluye la transcripción de la interacción, los archivos de audio de los participantes y las características faciales [83, p. 1].



Gráfica 13 - Descripción de los datos DAIC-WOZ [83, p. 1]

En la Gráfica 13 se encuentra el desglose de la Base de datos donde se encuentran 189 archivos .zip cada una contiene una carpeta de sesiones que va numerada desde 300- 492, también se tiene los siguientes archivos:

- train_split_Depression_AVEC2017.csv: Este archivo incluye:
  - ID de participantes
  - Etiquetas binarias PHQ8 (PHQ8 Scores>= 10)
  - PHQ8 Scores (Puntaje PHQ8)
  - Género de los participantes
  - Respuestas individuales para cada pregunta del cuestionario PHQ8 para la división oficial de train

  Posee la etiqueta de depresivo (1) y no depresivo (0) para 107 participantes de los 189 Participantes totales.

- dev_split_Depression_AVEC2017.csv: Este archivo incluye
  - ID de participantes
  - Etiquetas binarias PHQ8
  - Puntajes de PHQ8
  - Género de los participantes

- Respuestas individuales para cada pregunta del cuestionario PHQ8 para la división oficial de desarrollo.

Posee la etiqueta de depresivo (1) y no depresivo (0) para 35 participantes de los 189 Participantes totales.

- test_split_Depression_AVEC2017.csv: Este archivo comprende:
  - Los ID de los participantes.
  - El sexo del participante para la división oficial de prueba.

- full_test_split.csv: Este archivo comprende:
  - Los ID de los participantes.
  - Etiquetas binarias PHQ8 (PHQ8 Scores >= 10).
  - PHQ8 Scores.
  - El género de los participantes para la división oficial de full test.

Posee la etiqueta de depresivo (1) y no depresivo (0) para 47 participantes de los 189 Participantes totales.

8.1.2.2. Conjuntos de Datos de la Base de Datos DAIC-WOZ

Dentro de la base de Datos DAIC-WOZ del conjunto de Datos recolectado de 189 conversaciones de participantes, de las cuales se encuentran separadas en tres conjuntos diferentes los cuales son: train, dev, test los cuales tienen las siguientes distribuciones dentro de la base de datos:

- Train: De los 189 participantes, 107 de ellos hacen parte de train, dando un porcentaje de 56.61%.

  El conjunto de datos de entrenamiento se emplea de manera iterativa para aprender los parámetros del modelo, en el proceso de capacitación de un

modelo, el conjunto de entrenamiento tiene como propósito tomar una decisión sobre qué parámetros elegir dadas las enormes opciones para elegir dentro de un modelo [84, p. 5].

- Dev: De los 189 participantes, 35 de ellos hacen parte de dev, dando un porcentaje de 18.52%.

  El conjunto de datos de desarrollo se emplea en el caso de que se tenga varios modelos, teniendo como objetivo de este conjunto de datos clasificar los modelos en términos de su precisión y ayudar a decidir con qué modelo seguir adelante [84, p. 5].

- Test: De los 189 participantes, 47 de ellos hacen parte de test, dando un porcentaje de 24.87%.

  Al haber pasado los diferentes modelos o el modelo por los conjuntos de train y dev, se obtiene como resultado el mejor modelo, al cual se le pasa el conjunto de datos de test para confirmar si el modelo obtenido es un buen modelo o no, y cuánta precisión se puede obtener del mismo una vez que se lo implementa en el mundo real [84, p. 5].

Teniendo un porcentaje de distribución del total de 189 conversaciones para Train (56.61%), para Dev (18.52%), para Test (24.87%), donde es recomendable si se posee una cantidad muy pequeña de datos, intentar usar la mayor cantidad de datos posible para Train y hacer una división del orden 70/20/10 (train, dev, test) respectivamente [85], aunque los porcentajes como es el caso de la base de datos DAIC-WOZ pueden variar.

8.1.3. Análisis de PHQ-8 y cómo influye al detectar la depresión

8.1.3.1. PHQ-8: Descripción General

El PHQ-8 (Patient Health Questionnaire-8) es uno de los instrumentos que ha alcanzado mayor reconocimiento a

nivel internacional para el reconocimiento de la Depresión basados en los criterios del DSM-IV[19].

Es un inventario de autoinforme de opción múltiple, que se utiliza como una herramienta de detección y diagnóstico para los trastornos de salud mental como la depresión, la ansiedad, el alcohol, la alimentación y los trastornos somáticos.

La escala de depresión del cuestionario de salud del paciente de ocho ítems (PHQ-8) se estableció como una medida válida de diagnóstico y gravedad para los trastornos depresivos en grandes estudios clínicos [22].

Las puntuaciones de PHQ-8 se califican con una escala Likert[20] que va:

- de 0 (nunca).
- a 1 (varios días).
- 2 (más de la mitad de los días).
- y 3 (casi todos los días).

Por lo que el puntaje total va en un rango de 0 a 24. La gravedad de los síntomas puede organizarse en 4 categorías:

- 0-4 (mínimo)
- 5-9 (leve)
- 10-14 (moderado)
- 15-19 (moderadamente grave)
- 20-24 (grave o aguda)

Tabla 11 - Categorías de Depresión en base al puntaje de PHQ-8

Total Score (PHQ8)	Tipo de Categoría de Depresión
0-4	Depresión Mínima
5-9	Depresión Leve
10-14	Depresión Moderada

[19] El DSM-IV [106] incluye criterios para hacer el diagnóstico de un trastorno depresivo.
[20] Escala de Likert es una herramienta de medición que, a diferencia de preguntas con respuesta sí/no, permite medir actitudes y conocer el grado de conformidad de un encuestado con cualquier afirmación que se le proponga, resultando especialmente útil emplearla en situaciones en las que desea que la persona matice su opinión [19].

15-19	Depresión Moderadamente Grave
20-24	Depresión Grave o Aguda

El PHQ-8 se desarrolló como una herramienta de tamizaje[21], siendo los puntajes de corte recomendados entre 8 y 11 para un probable caso de depresión.

Para interpretar la puntuación obtenida en el cuestionario PHQ8 por parte del personal de salud idóneo, se emplea la Gráfica 14.

Puntuación	Acción
≤4	La puntuación indica que, probablemente, el paciente no necesita tratamiento para la depresión.
> 5 a 14	El médico debe utilizar su juicio clínico sobre el tratamiento, tomando en consideración la duración de los síntomas del paciente y su trastorno funcional.
≥15	Se justifica el tratamiento de la depresión con antidepresivos, psicoterapia o una combinación de tratamientos.

Gráfica 14 - Puntuación formulario PHQ-8 [86, p. 3]

8.1.3.2. PHQ-8: Formato de formulario

PHQ-8, consta de 8 preguntas son varios tipos de respuestas como se observa en la Gráfica 15:

[21] Herramienta de Tamizaje: Es aquella con un valor predictivo positivo elevado, que permite identificar a los individuos que realmente tienen un caso de depresión, son mediciones para establecer quién puede padecer cierta enfermedad y quien no, en cualquier momento de la vida [107].

PHQ-8

Over the last 2 weeks, how often have you been bothered by any of the following problems? (Use "✓" to indicate your answer)	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down	0	1	2	3
7. Troubleconcentratingonthings,suchas reading thenewspaperorwatchingtelevision.	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3

(For office coding: Total Score _____ = ___ + ___ + ___)

From the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MDPHQ). The PHQ was developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues. For research in formation, contact Dr. Spitzera trls8@columbia.edu. PRIME-MD® is a trademark of Pfizer Inc. Copyright© 1999 Pfizer Inc. All rights reserved. Reproduced with permission

Gráfica 15 - Estructura del Formulario de PHQ-8 [22, p. 167]

8.1.4. Establecer las características prosódicas que se usaran como predictores prometedores de la depresión

Las Características prosódicas son relevantes para la expresión de emociones por la combinación de los elementos que posee [62, p. 34], por esta razón se estable las características prosódicas como lo son:

- El tono
- El ritmo de las oraciones
- La acentuación
- La calidad de voz
- La articulación
- La entonación
- La longitud

Se van a tener como predictores prometedores de la depresión, haciendo alusión a elementos que tienen que ver con rasgos de sonido y tonos.

Estas características se manifiestan en las palabras para analizar su acentuación y entonación general en la oración o frase. Estos elementos son importantes tanto para la organización del discurso como para la expresión de emociones.

La depresión posee ciertas particularidades en el lenguaje y habla de una persona, donde se puede dejar oír como la enfermedad se presenta en las palabras de una persona que la padece, incluso cuando no reconoce estar depresivo, una persona puede mostrar señales de la enfermedad a través de la forma como describe acontecimientos, atribuyéndoles un significado propio de que posee depresión [41, p. 70].

8.1.5. Verificar la calidad de los archivos de Audio

Una parte de los datos que contiene DAIC-WOZ son del análisis de audio hechos con COVAREP[22] y transcripciones textuales de la entrevista, todas en un paquete de 189 archivos .zip donde cada uno contiene una carpeta de sesiones que va enumerada 300-492, dentro de cada carpeta los archivos de audio se encuentran en formato .wav.

8.1.5.1. Desglose de la estructura de la carpeta que contiene los archivos de audio

Cada carpeta tiene los siguientes archivos (donde XXX es el número de sesión, por ejemplo: XXX = 301 en la carpeta 301_P) así:

[22] COVAREP: Un repositorio de código abierto de algoritmos avanzados de procesamiento de voz [5].

XXX_P\

XXX_CLNF_features.txt
XXX_ CLNF _features3D.txt
XXX_ CLNF _gaze.txt
XXX_ CLNF _hog.bin
XXX_ CLNF _pose.txt
XXX_ CLNF _AUs.csv
XXX_AUDIO.wav
XXX_COVAREP.csv
XXX_FORMANT.csv
XXX_TRANSCRIPT.csv

Gráfica 16 - Descripción de la Carpeta – Entrevistas [83, p. 2]

En la Gráfica 16 se encuentra el desglose de la carpeta contenedora del audio, teniendo dentro de ella los siguientes archivos, de los cuales solo especificaremos los archivos de sonido y transcript, que son relevantes para el proyecto:

- XXX_AUDIO.wav
  Este archivo comprende una grabación de audio de una sesión, la grabación se realizó con un Sennheiser HSP 4-EW-3 a 16kHz ubicado o montado en la cabeza, ver Gráfica 17.

Gráfica 17 - Micrófono HSP 4-EW-3 [87]

- XXX_TRANSCRIPT.csv
  Los archivos de trascripción son archivos "separados por tabulaciones". Contienen todo lo hablado durante cada entrevista, tanto la parte de Ellie "entrevistador virtual" como la del participante o entrevistado.

8.1.5.2. Calidad de los archivos de audio

- Se excluyeron ciertas sesiones las cuales son: 342,394,398,460.

- Además, algunas sesiones fueron incluidas con notas especiales como lo son:

  - 373_P: Hay una interrupción alrededor de los minutos 5:52-7:00, un aliado ingresa a la sala para solucionar un problema técnico menor, la sesión continua y se completa.

  - 444_P: Hay una interrupción alrededor de los minutos 4:46-6:27, suena el teléfono del participante y un aliado entra a la sala para ayudar a apagarlo. La sesión continua y se completa.

  - 451_P, 458_P, 480_P: Las sesiones están técnicamente completas, pero faltan parte de las transcripciones de Ellie (la entrevistadora virtual). Las transcripciones de los participantes todavía se incluyen, pero sin las preguntas del entrevistador.

  - 402 _P: La grabación se corta 2 minutos antes del final de la conversación.

- Los archivos de audio restantes, pueden presentar ruidos o interferencias en algunos momentos durante el transcurso de la entrevista, teniendo:
  - Crujidos provenientes de los micrófonos.

- Ruidos de fondo como el murmullo del viento o un zumbido de línea de potencia.

- Interferencia de la Señal de Audio.

Dichos ruidos son señales no deseadas que se han mezclado con la señal útil que se quiere para el proyecto, siendo el resultado de diversos tipos de perturbaciones que tiende a enmascarar la información de la conversación llevada a cabo.

Los archivos .wav, se encuentran útiles, pero se deben pasar por ciertos filtros para eliminar dichos ruidos y obtener la señal más idónea posible para el estudio de la depresión.

8.2. Realizar el procesamiento de los datos con base de la metodología utilizada en el modelo DepressionDetect

8.2.1. Entorno de Trabajo

El proyecto fue realizado en un computador Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz de Arquitectura x64, con 2 núcleos, 4 procesos lógicos, L1 caché de 128kB, L2 caché de 512kB, L3 caché de 3 MB [88] y una configuración de memoria RAM de 12 GB.

Para el despliegue del proyecto se empleó el lenguaje de programación python acompañado de la tecnología Anaconda en su versión 4.7.12, durante todas las fases de implementación del proyecto, creando un entorno virtual acompañado de la versión de python 3.7.5, junto con pip, un administrador de paquetes para la instalación de librerías y aplicaciones secundarias desarrolladas en python, usadas durante el proyecto.

Se eligió el lenguaje de programación python por su legibilidad al momento de escribir o entender el código utilizado para la realización del proyecto, también por su gran variedad de librerías que son de gran ayuda.

Además, se utilizó la Librería Librosa en su versión 0.7.2 y Pandas en su versión 0.25.3 durante la fase de procesamiento

de los datos y la librería de Keras en su versión 2.4.2 acompañada con el framework TensorFlow en su versión 2.2.0 para el entrenamiento de la red convolucional, que permite facilitar el proceso de experimentación rápida y permite la creación de prototipos fácilmente.

- Librosa**:** Es un paquete de Python para análisis de audio y música. Proporciona los componentes básicos necesarios para crear sistemas de recuperación de información musical [18].

- Pandas: Es una herramienta de manipulación de datos de alto nivel, construida con el paquete Numpy y su estructura de datos clave es llamada el DataFrame [21], una clase de objetos especiales en el lenguaje de programación Python, denominados trama de datos, los cuales permiten almacenar y manipular datos tabulados en filas de observaciones y columnas de variables. [8].

- Keras: Es una API de redes neuronales de alto nivel, escrita en Python y capaz de correr sobre frameworks como TensorFlow, CNTK, o Theano.

- TensorFlow: Es una plataforma de código abierto de extremo a extremo para el aprendizaje automático, cuenta con un ecosistema integral y flexible de herramientas, bibliotecas y recursos de la comunidad que les permite a los investigadores impulsar un aprendizaje automático innovador [25]. Su arquitectura flexible hace que pueda ser ejecutada en multitud de plataformas (GPUs, CPUs o TPUs) y en ordenadores, clústeres de servidores e incluso dispositivos móviles [63].

8.2.2. Análisis previo de los audios dispuestos por la Base de Datos DAIC-WOZ

Ahora bien, con la base de datos descargada y las características prosódicas seleccionadas, segmentamos el discurso de la persona desde el silencio, otros hablantes y el ruido exterior, para ello se realizó primeramente un análisis de las Transcripciones de las entrevistas.

La base de datos DAIC-WOZ, en cada carpeta de cada participante cuenta con varios archivos de los cuales, se emplearon dos de ellos que son el archivo .wav que contiene el audio de la entrevista y el archivo .csv que contiene la transcripción de la entrevista, en base a eso y entendiendo que dichos archivos contienen todo lo hablado durante la entrevista, tanto la parte de Ellie "entrevistador virtual" como la del participante o entrevistado.

Al enfocarse en el entrevistado, durante la entrevista, la parte de la entrevistadora virtual Ellie, se despreció y se depuro, dejando solo la parte del entrevistado, para ello se utilizó la librería de python llamada Pandas para el tratamiento de los archivos .csv como DataFrames, depurando la parte del entrevistadora Ellie, donde se hizo una iteración en el directorio de la Base de Datos que contiene los 189 archivos .csv con las transcripciones de cada participante, extrayendo la parte donde solo habla el participante omitiendo a la entrevistadora virtual Ellie además se extrajo el rango de muestras donde habla solo el participante ver Tabla 12, para posteriormente editar el archivo .wav, en base al rango de muestras obtenidas de cada participante.

Tabla 12 - Rango de Muestras de Audio – Participante 300

dstart_sample	dstop_sample
1374332	1393075
1520965	1549850
1654367	1722722
….	…
13633691	13650670
13682863	13698959

Para la realización de lo anterior, se empleó un código realizado en Python en un notebook de Jupyter Notebook llamado Analysis_All_Transcripts.ipynb, dejando la información resultante en dos carpetas, una de ellas llamada transcripts_participants, la cual contiene los archivos .csv con

las transcripciones depuradas donde solo se encuentra los diálogos de cada participante y file_sample, la cual contiene los archivos .csv con los rangos de muestras donde habla el participante durante la entrevista.

8.2.3. Segmentación del Audio de los Participantes de la Base de Datos

Para la segmentación del audio se utilizó un DataFrame para almacenar la información de todas las conversaciones de los archivos .wav donde solo habla el participante, para ello cargamos los archivos .csv de los rangos de las muestras de audio donde solo habla el participante y cargamos los archivos originales .wav haciendo uso de la librería Librosa.

La cual tiene una función de carga que lee la ruta del archivo de audio y devuelve una tupla con dos elementos. El primer elemento es una 'serie temporal de audio' (tipo: matriz) correspondiente a la pista de audio. El segundo elemento de la tupla es la frecuencia de muestreo que se utiliza para procesar el audio, la frecuencia de muestreo predeterminada utilizada por Librosa es 22050 muestras por segundo [89].

Se hizo una iteración en las carpetas contenedoras de los archivos, al tener los archivos cargados para cada participante, se compararon los nombres de dichos archivos, al coincidir sus nombres, se fusionó en un solo DataFrame ambas informaciones, en donde, se unió en una matriz np.array de python, cada rango de muestras para cada participante quedando de esa forma el audio de dicho participante solo con la voz del participante, depurando completamente a la entrevistadora virtual Ellie dentro de la entrevista.

Al tener la matriz np.array completada para cada participante, se guardó en el DataFrame dicha información quedando con la siguiente estructura como se puede observar en la Tabla 13:

Tabla 13 - Dataframe – Segmentación del Audio de los Participantes

	participant	sample	audio_sample
0	300	3434506	[0.0004416694864630699, -0.0001553866168251261...

1	301	10483453	[-0.001894154935143888, -0.001956888008862734,...
2	302	4606979	[0.013585681095719337, 0.01482345536351204, 0....
3	303	14176606	[0.018198242411017418, 0.019573194906115532, 0...
4	304	7995327	[0.0010568039724603295, 0.0011041408870369196,...
...	...	...	...
184	488	9315900	[0.0016439873725175858, 0.0015722919488325715,...
185	489	3722264	[0.0009191472781822085, 0.0008580769062973559,...
186	490	4099096	[-9.321090328739956e-05, -0.000206125841941684...
187	491	9119442	[-0.00012431594950612634, -0.00012830703053623...
188	492	10500650	[0.0017372781876474619, 0.0017825436079874635,..

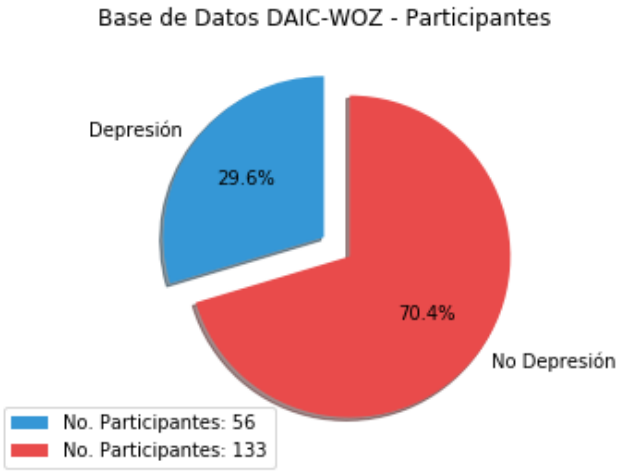
Al tener dicha información en el Dataframe, se procedió a guardarla localmente como un archivo .wav, ya que para hacer la segmentación se trataron los datos como una matriz, para ello se empleó una función de salida de la librería Librosa para exportar la información de la entrevista de cada participante como un archivo .wav

Para la realización de lo anterior, se empleó un código realizado en Python en un notebook de Jupyter Notebook llamado Create_Sample_Audio.ipynb, dejando la información resultante en una carpeta, llamada `audio_participant`, la cual contiene los archivos .wav con el siguiente formato XXX_AUDIO_PARTICIPANT.wav, donde XXX hace alusión al número del participante.

8.2.4. Desequilibrio de Clase dentro de la Base de Datos

8.2.4.1. Desbalanceo de la Base de Datos

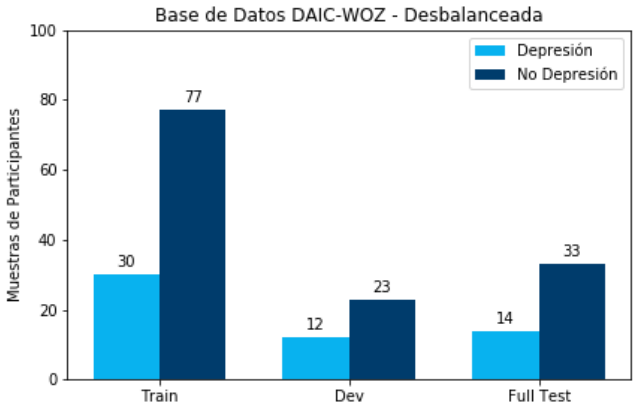
La Base de Datos cuenta con 189 participantes, de los cuales se tiene una proporción dispareja de aquellos que poseen y los que no poseen depresión, ilustrado en la Gráfica 18:



Gráfica 18 - Base de Datos DAIC-WOZ - Participantes Desbalanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

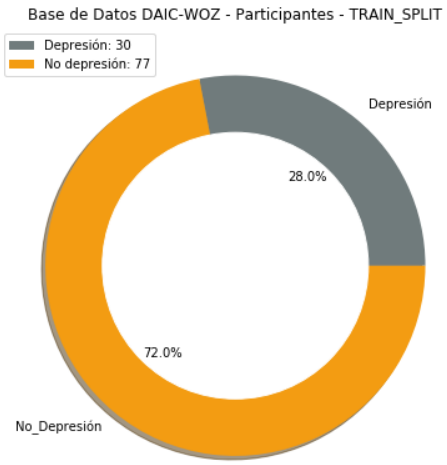
Dentro de la base de datos también existe algunas muestras o clases desproporcionadas respecto a las otras, donde hay más o menos cantidad de información respecto a las demás, esto provoca un desbalanceo en los datos que se desean utilizar para el entrenamiento de la red, la base de datos posee los conjuntos de datos de train, dev, test, los cuales poseen la siguiente información, como se observa en la Gráfica 19:



Gráfica 19 - Base de Datos DAIC-WOZ - Desbalanceada

Fuente: Autoría propia, Edward Camilo Villota Taramuel

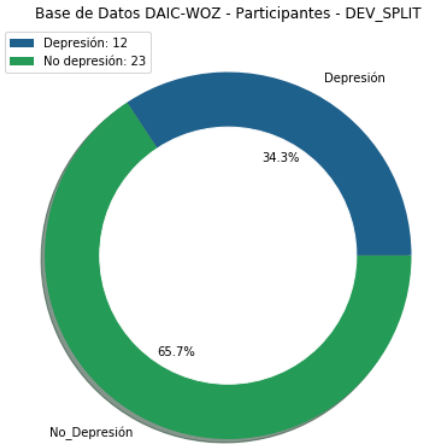
- Train: El conjunto de datos de Train presenta un desbalanceo en su conjunto de datos, con una relación (28.0%, 72.0%), Participantes con Depresión y Participantes sin Depresión respectivamente, ilustrado en la Gráfica 20.



Gráfica 20 - Base de Datos DAIC-WOZ - Participantes – TRAIN Desbalanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

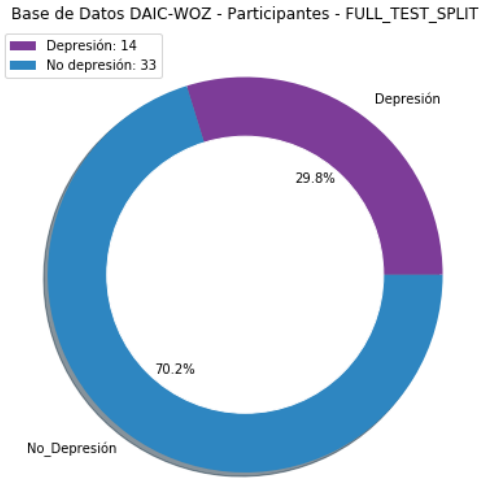
- Dev: El conjunto de datos de Dev presenta un desbalanceo en su conjunto de datos, con una relación (34.3%, 65.7%), Participantes con Depresión y Participantes sin Depresión respectivamente, ilustrado en la Gráfica 21.



Gráfica 21 - Base de Datos DAIC-WOZ - Participantes - DEV Desbalanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

- Test: El conjunto de datos de Test presenta un desbalanceo en su conjunto de datos, con una relación (29.8%, 70.2%), Participantes con Depresión y Participantes sin Depresión respectivamente, ilustrado en la Gráfica 22.



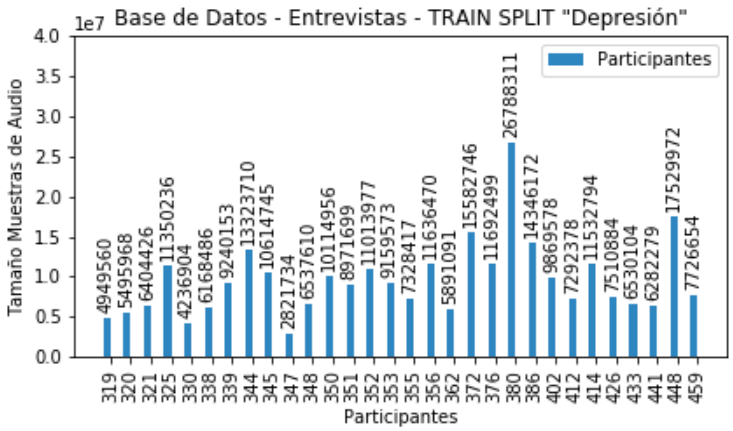
Gráfica 22 - Base de Datos DAIC-WOZ - Participantes - TEST Desbalanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

8.2.4.2. Desbalanceo de la Muestras de Audio de los Participantes

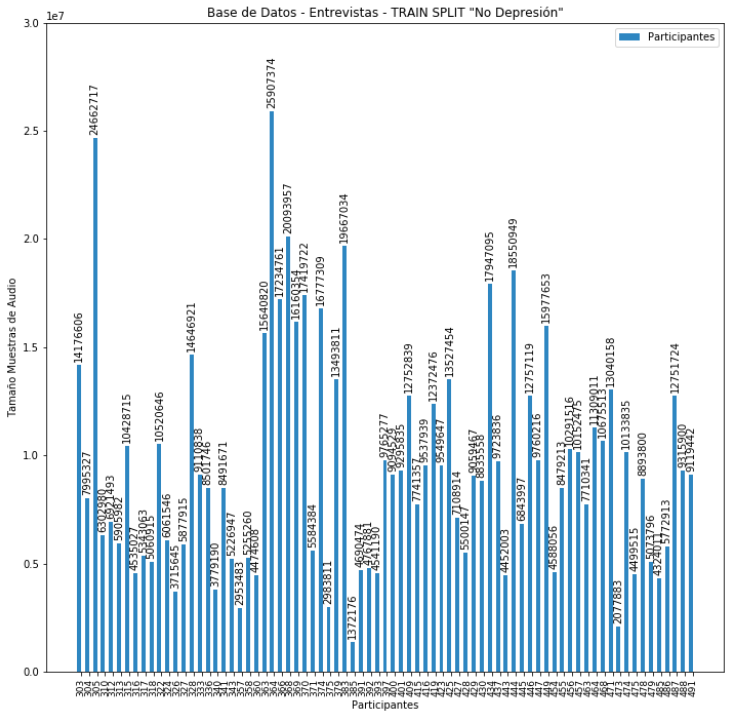
Dentro de la base de datos las muestras de audio se encuentran desproporcionadas respecto a otras, donde hay más o menos cantidad de información respecto a las demás, en relación a la duración de la entrevista para cada participante, esto provoca un desbalanceo en los datos de las muestras de audio que se desean utilizar para el entrenamiento de la red, dentro de los conjuntos de datos de train, dev, test, se posee la siguiente información, tanto para los participantes depresivos, como para los participantes que no tienen depresión, como podemos observar a continuación:

- Train: El conjunto de datos de las muestras de audio de los participantes de train presenta un desbalanceo en su conjunto de datos, donde hay una brecha demasiado grande en relación a la entrevista con menor duración y la entrevista con mayor duración, tanto para el conjunto de Train Depresivo, como para el conjunto de Train No Depresivo teniendo una relación (2.821.734, 26.788.311) y (1.372.176, 25.907.374) respectivamente, ver Gráfica 23 y Gráfica 24.



Gráfica 23 - Muestras de Audio - Entrevistas - TRAIN "Depresión" Desbalanceado

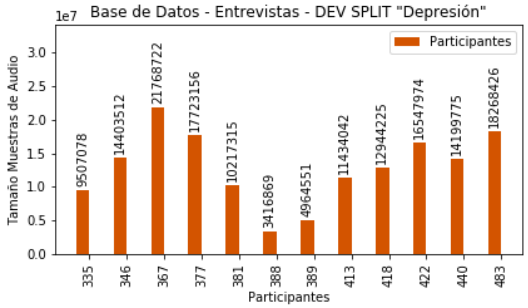
Fuente: Autoría propia, Edward Camilo Villota Taramuel



Gráfica 24- Muestras de Audio - Entrevistas - TRAIN "No Depresión" Desbalanceado

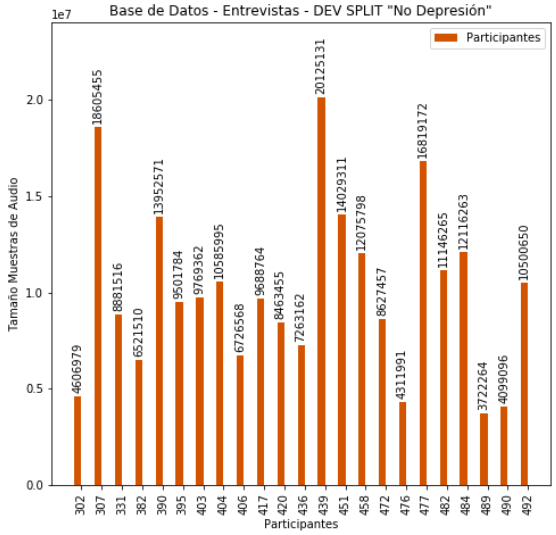
Fuente: Autoría propia, Edward Camilo Villota Taramuel

- Dev: El conjunto de datos de las Muestras de Audio de los Participantes de Dev presenta un desbalanceo en su conjunto de datos, donde hay una brecha demasiado grande en relación a la entrevista con menor duración y la entrevista con mayor duración, tanto para el conjunto de Dev Depresivo, como para el conjunto de Dev No Depresivo teniendo una relación (3.416.869,21.768.722) y (3.722.264,20.125.131) respectivamente, ver Gráfica 25 y Gráfica 26.



Gráfica 25 - Muestras de Audio - Entrevistas - DEV "Depresión" Desbalanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

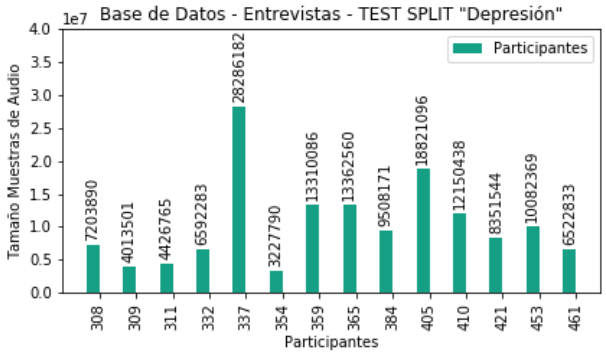


Gráfica 26 - Muestras de Audio - Entrevistas - DEV "No Depresión" Desbalanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

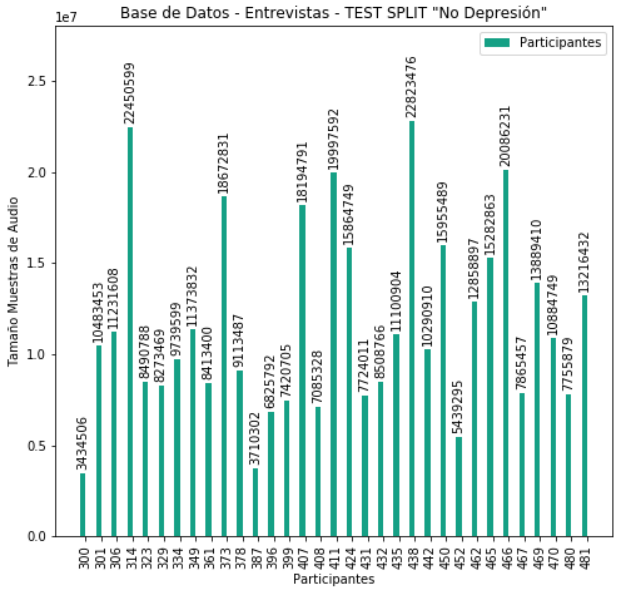
- Test: El conjunto de datos de las muestras de audio de los participantes de test presenta un desbalanceo en su conjunto de datos, donde hay una brecha demasiado grande en relación a la entrevista con menor duración y la entrevista con mayor duración, tanto para el conjunto de Test Depresivo, como para el conjunto de Test No Depresivo teniendo una relación

(3.227.790,28.286.182) y (3.434.506,22.823.476) respectivamente, ver Gráfica 27 y Gráfica 28.



Gráfica 27 - Muestras de Audio - Entrevistas - TEST "Depresión" Desbalanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel



Gráfica 28 - Muestras de Audio - Entrevistas - TEST "No Depresión" Desbalanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

8.2.5. Balanceo de la Base de Datos sin Perdida de Información

Se Realizó una serie de procedimientos para solucionar el desequilibrio de clase existente en la Base de Datos y el desbalanceo de las muestras de audio de los participantes.

Para ello se hizo uso de los audios .wav después de la segmentación, dichos archivos se cargaron mediante la librería Librosa y posteriormente fueron puestos en un DataFrame para almacenarlos, ver Tabla 13.

Se añadió al Dataframe, el valor Binario de Depresión para cada participante obtenido de los archivos .csv de train, dev y full test de la base de datos DAIC-WOZ, los cuales poseen la etiqueta binaria de depresivo “1”, no depresivo “0” para cada participante de cada conjunto de datos, para ello se creó una nueva columna llamada phq8_binary en el DataFrame.

Se convirtieron los archivos .csv de train, dev, full test en un DataFrame respectivamente para cada uno de ellos, se recorrió el DataFrame principal y cada uno de los otros DataFrame (train, dev y full test), comparando el ID de los participantes y encontrando similitud entre ellos, uniendo posteriormente el valor de la columna que posee la etiqueta binaria de los DataFrame (train, dev, full test), en el DataFrame principal en la columna phq8_binary.

Quedando como resultado el Dataframe Principal como se puede observar en Tabla 14:

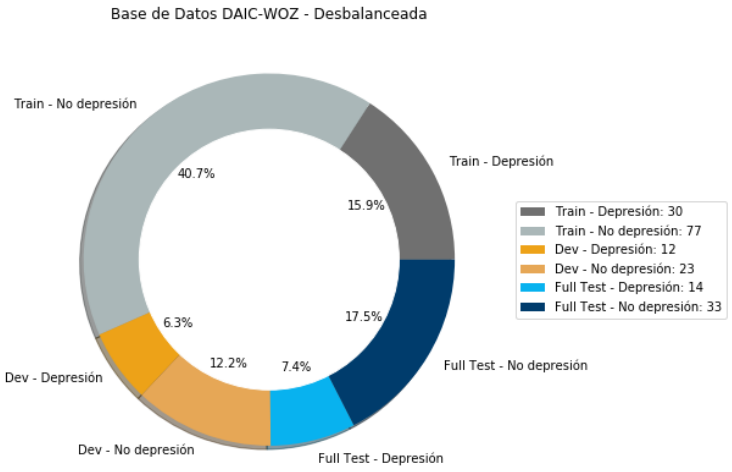
Tabla 14 - Dataframe – Etiqueta Binaria "Depresión, No Depresión" de los Participantes

	participant	phq8_binary	sample	audio_sample
0	300	0	3434506	[0.0004416695, -0.00015538662, 0.00036901518, ...
1	301	0	10483453	[-0.0018941549, -0.001956888, -0.0024183441, -...
2	302	0	4606979	[0.013585681, 0.014823455, 0.01562177, 0.01621...
3	303	0	14176606	[0.018198242, 0.019573195, 0.020426963, 0.0203...
4	304	0	7995327	[0.001056804, 0.0011041409, 0.0015114165, 0.00...
...	...	...	...	...

184	488	0	9315900	[0.0016439874, 0.001572292, 0.0013316689, 0.00...
185	489	0	3722264	[0.0009191473, 0.0008580769, 0.0008644862, 0.0...
186	490	0	4099096	[-9.32109e-05, -0.00020612584, -0.00020311444,...
187	491	0	9119442	[-0.00012431595, -0.00012830703, -0.0001888585...
188	492	0	10500650	[0.0017372782, 0.0017825436, 0.0017300711, 0.0...

8.2.5.1. Separación del Conjunto de Datos General en diferentes conjuntos (Train, Dev, Full Test)

Se tiene el Dataframe Principal ver Tabla 14, que contiene toda la información de los participantes. Se separó la información contenida en dicho DataFrame en diferentes Dataframe dependiendo a que grupo (train, dev, full test) pertenece dicho participante y si el participante tiene o no etiqueta depresiva, ilustrado en la Gráfica 29.



Gráfica 29 - Conjuntos de Datos de los Participantes con Etiqueta Binaria "Depresión, No Depresión"

Fuente: Autoría propia, Edward Camilo Villota Taramuel

- Train – Depresivo: Contiene todos los participantes que tienen depresión del conjunto de entrenamiento (train), ver Tabla 15

Tabla 15 - Dataframe – Etiqueta Binaria "Depresión" de los Participantes -TRAIN

	participant	phq8_binary	sample	audio_sample
19	319	1	4949560	[0.0001386485, 9.320122e-05, 0.00017489721, 2....
20	320	1	5495968	[0.0002724654, 6.617844e-05, -6.5001266e-05, -...
…	…	…	…	…
156	459	1	7726654	[0.0019014471, 0.0019262932, 0.0019078627, 0.0...

De manera gráfica, ver Gráfica 23.

- Train – No Depresivo: Contiene todos los participantes que No tienen depresión del conjunto de entrenamiento (train), ver Tabla 16.

Tabla 16 - Dataframe – Etiqueta Binaria "No Depresión" de los Participantes -TRAIN

	participant	phq8_binary	sample	audio_sample
3	303	0	14176606	[0.018198242, 0.019573195, 0.020426963, 0.0203...
4	304	0	7995327	[0.001056804, 0.0011041409, 0.0015114165, 0.00...
…	…	…	…	…
187	491	0	9119442	[-0.00012431595, -0.00012830703, -0.0001888585...

De manera gráfica, ver Gráfica 24.

- Dev – Depresivo: Contiene todos los participantes que tienen depresión del conjunto de desarrollo (dev), ver Tabla 17.

Tabla 17 - Dataframe – Etiqueta Binaria "Depresión" de los Participantes -DEV

	participant	phq8_binary	sample	audio_sample
35	335	1	9507078	[-0.00043354864, -0.00038251665, -0.0004263618...

45	346	1	14403512	[0.00036809198, 0.00030944703, 0.00026962554, ...
…	…	…	…	…
179	483	1	18268426	[0.0012156497, 0.0015926503, 0.0018756543, 0.0...

De manera gráfica, ver Gráfica 25.

- Dev – No Depresivo: Contiene todos los participantes que No tienen depresión del conjunto de desarrollo (dev), ver Tabla 18.

Tabla 18 - Dataframe – Etiqueta Binaria "No Depresión" de los Participantes -DEV

	participant	phq8_binary	sample	audio_sample
2	302	0	4606979	[0.013585681, 0.014823455, 0.01562177, 0.01621...
7	307	0	18605455	[0.01097687, 0.010930862, 0.010409966, 0.00970...
…	…	…	…	…
188	492	0	10500650	[0.0017372782, 0.0017825436, 0.0017300711, 0.0...

De manera gráfica, ver Gráfica 26.

- Test – No Depresivo: Contiene todos los participantes que tienen depresión del conjunto de prueba (test), ver Tabla 19.

Tabla 19 - Dataframe – Etiqueta Binaria "Depresión" de los Participantes -TEST

	participant	phq8_binary	sample	audio_sample
8	308	1	7203890	[-0.01390958, -0.015919633, -0.017222118, -0.0...
9	309	1	4013501	[-0.0008186057, -0.0010105834, -0.0024073785, ...
…	…	…	…	…
157	461	1	6522833	[0.00043677646, 0.0003867479, 0.0003163492, 0....

De manera gráfica, ver Gráfica 27.

- Test – No Depresivo: Contiene todos los participantes que No tienen depresión del conjunto de prueba (test), ver Tabla 20.

Tabla 20 - Dataframe – Etiqueta Binaria "No Depresión" de los Participantes -TEST

	participant	phq8_binary	sample	audio_sample
0	300	0	3434506	[0.0004416695, -0.00015538662, 0.00036901518, ...
1	301	0	10483453	[-0.0018941549, -0.001956888, -0.0024183441, -...
…	…	…	…	…
177	481	0	13216432	[0.0014984695, 0.0015184081, 0.0015007398, 0.0...

De manera gráfica, ver Gráfica 28.

8.2.5.2. Subdivisión de muestras de Audio para Train, Dev y Test

Dentro de la base de datos las muestras de audio se encuentran desproporcionadas respecto a otras, donde hay más o menos cantidad de información respecto a las demás, en relación con la duración de la entrevista para cada participante, ilustrado en la Tabla 21.

Tabla 21 - Duración de la Entrevistas de los Participantes (Muestras de Audio)

Conjunto de Datos	Duración de la Entrevistas de los Participantes (Muestras de Audio)	
	Tamaño Muestra - Menor	Tamaño Muestra - Mayor
Train – Depresión	2.821.734	26.788.311
Train – No Depresión	1.372.176	25.907.374
Dev – Depresión	3.416.869	21.768.722
Dev – No Depresión	3.722.264	20.125.131
Test - Depresión	3.227.790	28.286.182

Test – No Depresión	3.434.506	22.823.476

Presentando un desbalanceo en las muestras de audio que se desean utilizar para el entrenamiento de la red, para ello se decidió submuestrear las muestras de audio acomodándolas entre un rango de 2.000.000 a 4.000.000 muestras, haciendo divisiones de las entrevistas de los participantes sin pérdida de información.

Se construyó una función de submuestreo en python donde se carga los archivos .csv con los rangos de las muestras de audio donde solo habla el participante, el audio original .wav de cada participante y se le manda rango mínimo de corte y rango máximo de corte, que este caso es [2.000.000,4.000.000], al enviar esta información, la función recorta o modifica el audio original en base a los rangos de las muestras de audio donde solo habla el participante, sacando submuestras para cada participante, si dicha muestra se encuentra entre el rango máximo de corte y el rango mínimo de corte.

Ejemplo de Ilustración:

Para el participante 301_AUDIO.wav, el cual tiene una duración de entrevista de 10.483.453 de muestras de audio, mediante la función de submuestreo implementada, se modifica dicho archivo de audio, creando submuestras de dicha entrevista teniendo como resultado cuatro archivos .wav:

- *301_1_AUDIO_PARTICIPANT.wav con tamaño de muestra de: 2.037.862*
- *301_2_AUDIO_PARTICIPANT.wav con tamaño de muestra de: 2.065.203*
- *301_3_AUDIO_PARTICIPANT.wav con tamaño de muestra de: 2.232.561*
- *301_4_AUDIO_PARTICIPANT.wav con tamaño de muestra de: 4.147.827)*

Realizando lo anterior, el número de participantes aumento y las muestras de audio se acercaron más entre

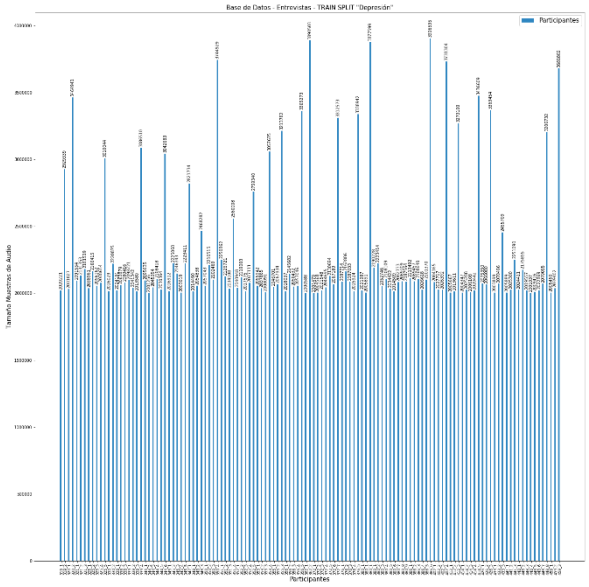
sí al hacer el submuestreo, solucionando parcialmente el problema entre las entrevistas muy largas y las entrevistas muy cortas de los participantes, quedando los datos de la siguiente manera, ver Tabla 22

Tabla 22 - Duración de la Entrevistas de los Participantes (Muestras de Audio) - Submuestreo

Conjunto de Datos	**Duración de la Entrevistas de los Participantes (Muestras de Audio) - Submuestreo**	
	Tamaño Muestra - Menor	**Tamaño Muestra - Mayor**
Train – Depresión	2.000.588	3.906.595
Train – No Depresión	1.372.176	4.107.477
Dev – Depresión	2.000.707	3.944.309
Dev – No Depresión	2.000.907	3.980.911
Test - Depresión	2.000.156	4.001.858
Test – No Depresión	2.000.377	4.147.827

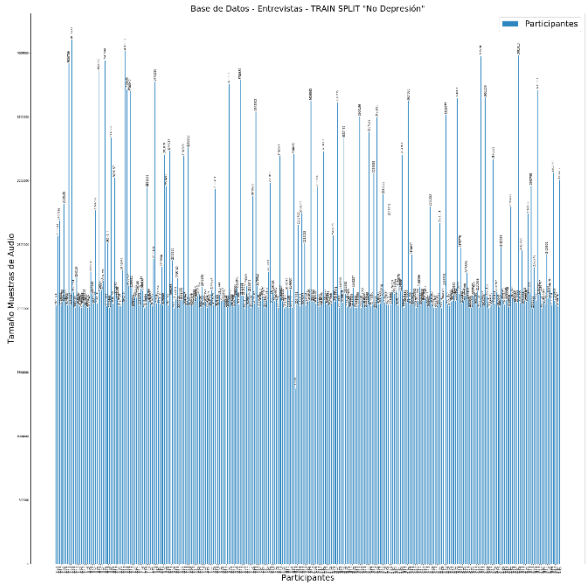
Dentro de los conjuntos de datos de train, dev, test, después de hacer el Submuestreo en las muestras de audio para los participantes de cada conjunto de datos se posee la siguiente información, tanto para los participantes depresivos, como para los participantes que no tienen depresión, como se observa a continuación:

- Train: El conjunto de datos de las Muestras de Audio de los Participantes de Train presenta para el conjunto de Train Depresivo, como para el conjunto de Train No Depresivo una relación (2.000.588,3.906.595) y (1.372.176,4.107.477) respectivamente, ilustrado en la Gráfica 30 y Gráfica 31.



Gráfica 30 - Muestras de Audio - Entrevistas - TRAIN "Depresión" Submuestreo

Fuente: Autoría propia, Edward Camilo Villota Taramuel

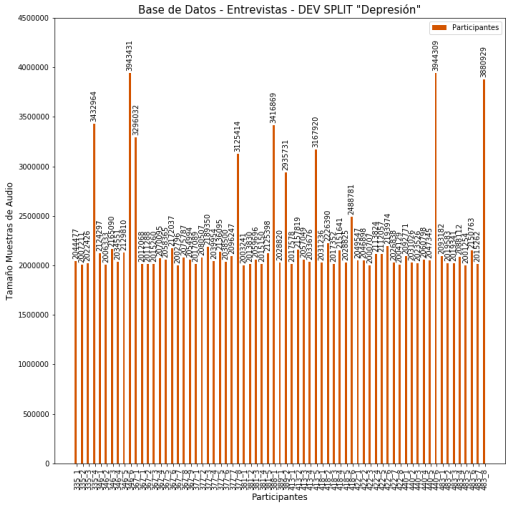


Gráfica 31 - Muestras de Audio - Entrevistas - TRAIN "No Depresión" Submuestreo

Fuente: Autoría propia, Edward Camilo Villota Taramuel

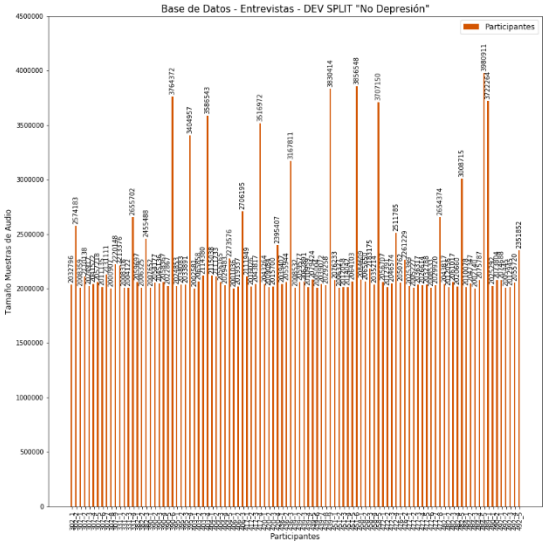
- Dev: El conjunto de datos de las Muestras de Audio de los Participantes de Dev presenta para el

conjunto de Dev Depresivo, como para el conjunto de Dev No Depresivo una relación (2.000.707,3.944.309) y (2.000.907,3.980.911) respectivamente, ilustrado en la Gráfica 32 y Gráfica 33.



Gráfica 32 - Muestras de Audio - Entrevistas - DEV "Depresión" Submuestreo

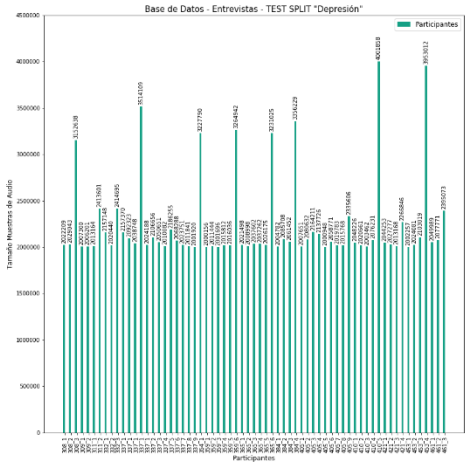
Fuente: Autoría propia, Edward Camilo Villota Taramuel



Gráfica 33 - Muestras de Audio - Entrevistas - DEV "No Depresión" Submuestreo

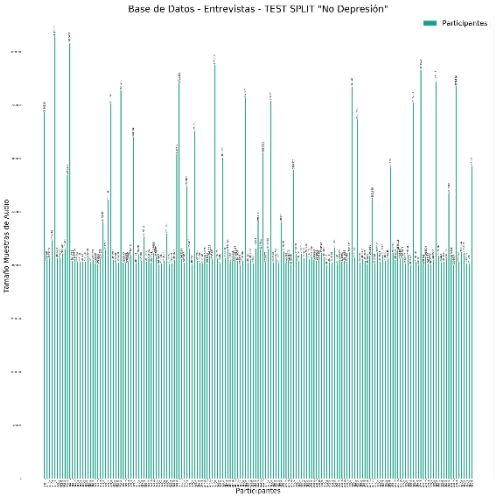
Fuente: Autoría propia, Edward Camilo Villota Taramuel

- Test: El conjunto de datos de las Muestras de Audio de los Participantes de Test presenta para el conjunto de Test Depresivo, como para el conjunto de Test No Depresivo una relación (2.000.156,4.001.858) y (2.000.377,4.147.827) respectivamente, ilustrado en la Gráfica 34 y Gráfica 35.



Gráfica 34 - Muestras de Audio - Entrevistas - TEST "Depresión" Submuestreo

Fuente: Autoría propia, Edward Camilo Villota Taramuel



Gráfica 35 - Muestras de Audio - Entrevistas - TEST "No Depresión" Submuestreo

Fuente: Autoría propia, Edward Camilo Villota Taramuel

Para la realización de lo anterior, se empleó un código realizado en Python en un notebook de Jupyter Notebook llamado Create_Partition_Audio.ipynb, dejando la información resultante en una serie de carpetas, llamadas `train_participant_1`, `train_participant_0`, `dev_participant_1`, `dev_participant_0`, `full_test_participant_1`, `full_test_participant_0`, las cuales contienen los archivos .wav con el siguiente formato XXX_X_AUDIO_PARTICIPANT.wav, donde XXX hace alusión a el número del participante y X hace alusión al número en las que fue dividida la entrevista de dicho participante.

8.2.5.3. Balanceo de la Base de Datos DAIC-WOZ

Se cargaron los archivos que se encuentran en las carpetas train_participant_1, train_participant_0, dev_participant_1, dev_participant_0, full_test_participant_1, full_test_participant_0 y se almacenó la información, en un DataFrame por cada carpeta, sabiendo también que el número de participantes aumentó, teniendo las siguientes cifras para cada conjunto de datos (train, dev, test), ver Tabla 23

Tabla 23 – Base de Datos DAIC-WOZ - Desbalanceado

Conjunto de Datos	Número de Participantes
Train – Depresión	125
Train – No Depresión	321
Dev – Depresión	69
Dev – No Depresión	103
Test - Depresión	65
Test – No Depresión	169

Donde se observa un desbalanceo en los datos, teniendo una proporción dispareja de aquellos que poseen y los que no poseen depresión, donde hay más o menos cantidad de información respecto a las demás.

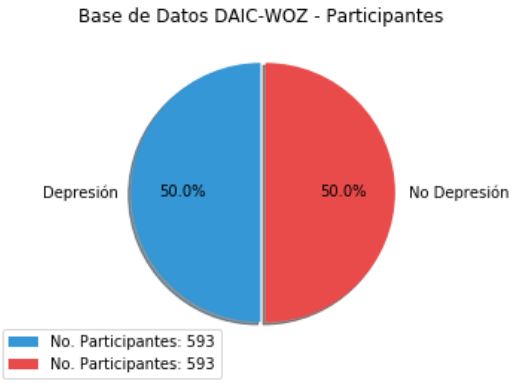
Se utilizó la librería sklearn de python, importando la función resample (`from sklearn.utlis import resample`) para muestrear los DataFrames de los conjuntos de datos de manera consistente, igualando el tamaño de cada conjunto de datos (train, dev, test) para balancear

el tamaño de muestras de los participantes en la Base de Datos.

```
sklearn.utlis.resample(*arrays,**options)
```

Esta función permite volver a muestrear matrices o matrices dispersas de tamaño consistente [90].

Una vez se aplicó la función de `resample()`, se solucionó el desequilibrio de clase para cada conjunto de datos de participantes quedando la base de datos DAIC-WOZ con datos en igual proporción, como se observa en la Gráfica 36.



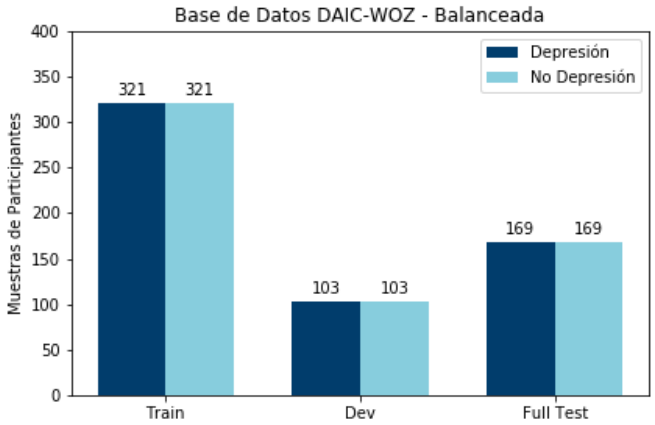
Gráfica 36 - Base de Datos DAIC-WOZ - Participantes Balanceado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

Quedando una proporción pareja de aquellos que poseen y los que no poseen depresión, para cada conjunto de datos, ver Tabla 24 y Gráfica 37.

Tabla 24 - Base de Datos DAIC-WOZ - Balanceado

Conjunto de Datos	Número de Participantes
Train – Depresión	321
Train – No Depresión	321
Dev – Depresión	103
Dev – No Depresión	103
Test - Depresión	169
Test – No Depresión	169



Gráfica 37 - Base de Datos DAIC-WOZ - Balanceada

Fuente: Autoría propia, Edward Camilo Villota Taramuel

Para la realización de lo anterior, se empleó un código realizado en Python en un notebook de Jupyter Notebook llamado Create_Resample_Audio.ipynb.

8.2.5.4. Filtros de Procesamiento de Audio para limpieza de las entrevistas de los participantes

Los archivos de audio .wav de la base de datos DAIC-WOZ, en su mayoría presentan ruidos o interferencias en algunos momentos durante el transcurso de la entrevista, como crujidos provenientes de los micrófonos, ruidos de fondo como el murmullo del viento o un zumbido de línea de potencia e interferencia de la Señal de Audio.

Dichos ruidos son señales no deseadas que se han mezclado con la señal útil que se quiere para el proyecto, siendo el resultado de diversos tipos de perturbaciones que tiende a enmascarar la información de la conversación llevada a cabo.

Se empleó para quitar los ruidos de la señal de audio de los archivos .wav, un Filtro Pasa Bajo (F.P.B) y un Filtro Rechaza Banda (F.R.Bd), haciendo uso de la librería Think DSP de python la cual es idónea para procesamiento de señal digital, que proporciona clases y funciones para trabajar con señales [91], obteniendo la señal más idónea posible para el estudio de la depresión.

- F.P.B: Filtro Pasa Baja es un filtro que permite el paso de frecuencias más bajas que una denominada frecuencia de corte y atenúa todos los componentes de frecuencia por encima de dicha frecuencia de corte [92, p. 7].

- F.R.Bd: Filtro Rechaza Banda es un filtro que posee una frecuencia de corte inferior y una frecuencia de corte superior, dejando solo pasar las frecuencias que no se encuentran entre esos rangos de Frecuencia de Corte, atenuando los componentes en la banda de frecuencias entre esos dos rangos [91, p. 7].

Se empleó un F.P.B con una Frecuencia de Corte (Fc) de 2.000 Hz, que dejo pasar las frecuencias menores a 2.000 Hz y un F.R.Bd que eliminó el ruido de banda agosta[23] que esta entre 50 Hz y 60 Hz.

Para ello, se instaló la librería Think DSP, en su versión 1.1.3, mediante el siguiente comando:

```
pip install thinkx
```

El cual posee una serie de funciones para trabajar con archivos .wav, teniendo la siguiente relación en su estructura para implementar Filtros de procesamiento de audio, ver Gráfica 38.

Gráfica 38 - Relaciones entre las clases de la Librería ThinkDSP [91, p. 8]

Se cargaron los archivos que se encuentran en las carpetas train_participant_1, train_participant_0, dev_participant_1, dev_participant_0, full_test_participant_1, full_test_participant_0, y se

[23] Ruido de Banda Agosta: Esta clase de ruido tiene un nivel y una frecuencia constantes, como el ruido de la línea de alimentación, el zumbido de un motor, entre otros, son ruidos de banda angosta normalmente provocados por cables con mala conexión a tierra y malos blindajes [109].

leyeron los archivos de audio mediante la función `thinkdsp.read_wave()`.

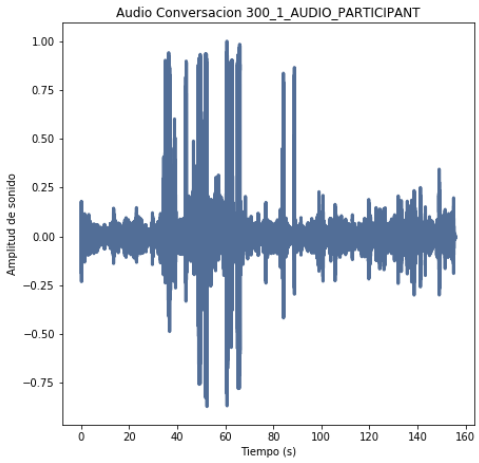
Se presentó un inconveniente ya que la función `thinkdsp.read_wave()` de la librería Think DSP no leyó los archivos .wav, para ello se usó un convertidor en línea [93] para convertir el formato de los archivos de audio .wav de las diferentes carpetas en un formato .wav compatible con la función de la librería Think DSP.

Se usó el siguiente formato para los archivos .wav convertidos:

- Frecuencia de muestreo: 22050 Hz
- Canales: 2 (Estéreo)

De esa manera la función de la librería Think DSP, acepto los archivos. Se cargaron los archivos .wav para cada participante de cada conjunto de datos (train, dev, test) mediante la función `thinkdsp.read_wave()`, la cual asigna a una variable `wave` el archivo de audio, cargándolo en el sistema.

Pudiendo así ver de manera visual, la gráfica de audio de cualquier participante como se puede ver en la Gráfica 39 de audio de un participante antes de filtrar la señal de audio:



Gráfica 39 - Audio Conversación - 300_1_AUDIO_PARTICIPANT - Sin Filtrar

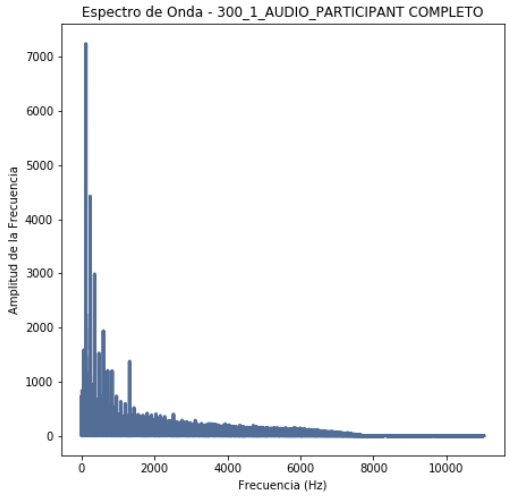
Fuente: Autoría propia, Edward Camilo Villota Taramuel

Pudiendo acceder a la variable `wave` mediante herramientas de la librería [91], como:

- `wave.ys`, retorna la 'serie temporal de audio' (tipo: matriz) correspondiente a la pista de audio.
- `wave.framerate` retorna la frecuencia de muestreo que se utiliza para procesar el audio, la frecuencia de muestreo predeterminada utilizada por la Librería Think DSP es 22050 muestras por segundo.

Los archivos de audio sin filtrar en el sistema se convirtieron en Espectros de Onda[24] mediante la función `wave.make_spectrum()`, haciendo uso de la DTF[25] para expresar la señal de audio como la suma de sinusoides con diferentes frecuencias.

Como se puede observar en la Gráfica 40 el espectro de onda de un participante antes de filtrar la señal de audio:



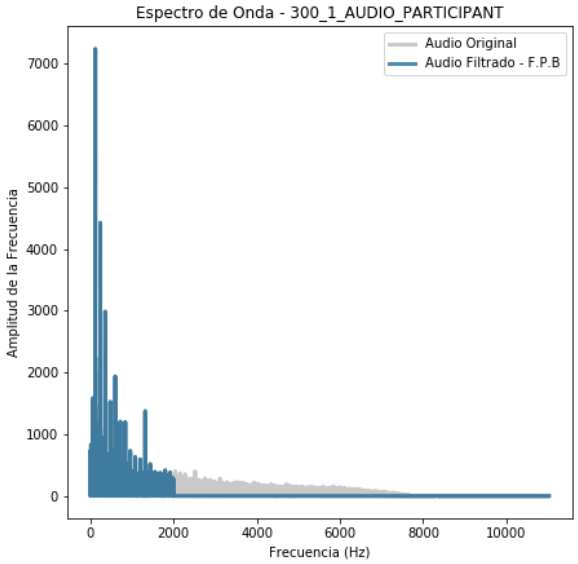
Gráfica 40 - Espectro de Onda - 300_1_AUDIO_PARTICIPANT - Sin Filtrar

Fuente: Autoría propia, Edward Camilo Villota Taramuel

[24] Espectro de Onda: Es la representación gráfica de una señal en función de frecuencia (x) vs amplitud (y) [110].

[25] DTF (Discrete Fourier Transform): Es un método que permite la transformación de una función del tiempo en una función de la frecuencia [91, p. 3].

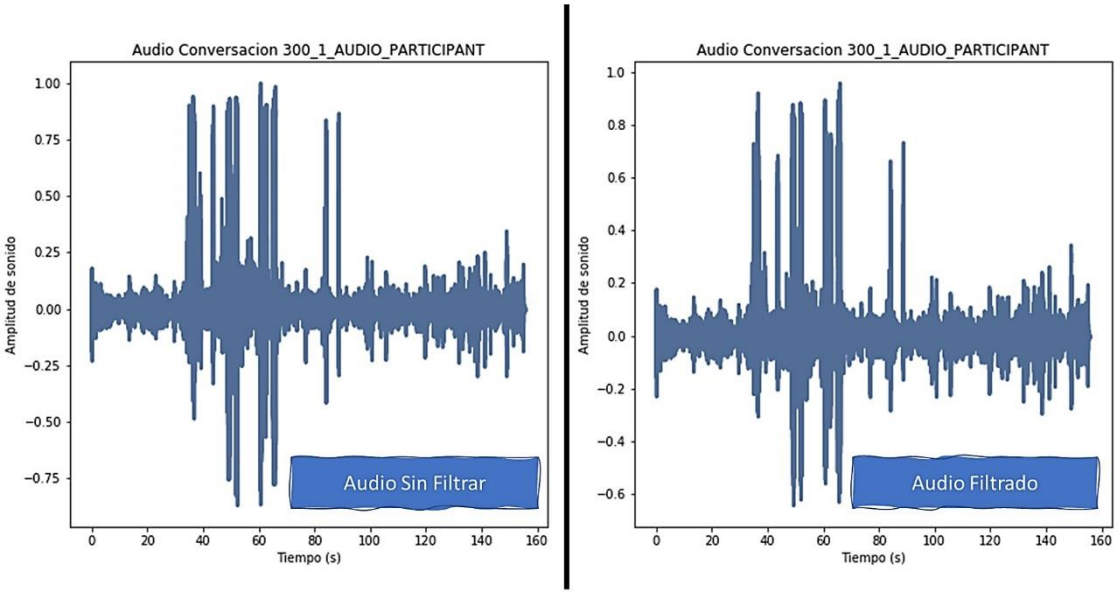
Con el Espectro de Onda de los participantes creado, se aplicó el filtro F.P.B con una frecuencia de corte (Fc) de 2.000 Hz, mediante la función `spectrum.low_pass()` y un F.R.Bd entre 50Hz y 60Hz mediante la función `spectrum.band_stop()`, filtrando así la información, una vez se filtró la información se procedió a regresar de un espectro de onda a un dato en formato wave, y posteriormente a una señal de audio .wav y se almacenó la información de los audios filtrados con el formato XXX_X_AUDIO_PARTICIPANT.wav, donde XXX hace alusión a el número del participante y X hace alusión al número en las que fue dividida la entrevista de dicho participante.



Gráfica 41 - Espectro de Onda - 300_1_AUDIO_PARTICIPANT - Filtrado F.P.B

Fuente: Autoría propia, Edward Camilo Villota Taramuel

Resultado de la señal de audio antes y después del filtrado se muestra en la Gráfica 42.



Gráfica 42 - Comparación Audio Conversación - 300_1_AUDIO_PARTICIPANT - Sin Filtrar y Filtrado

Fuente: Autoría propia, Edward Camilo Villota Taramuel

Para la realización de lo anterior, se empleó un código realizado en Python en un notebook de Jupyter Notebook llamado Create_Filters_Audio.ipynb, dejando la información resultante en una serie de carpetas, llamadas `train_participant_1`, `train_participant_0`, `dev_participant_1`, `dev_participant_0`, `full_test_participant_1`, `full_test_participant_0`, las cuales contienen los archivos .wav filtrados.

8.3. Establecer a partir de la Base de Datos DAIC-WOZ el conjunto de datos de entrenamiento y el conjunto de datos de prueba que garantice una información sin sesgo

8.3.1. Remuestreo de las muestras de Audio Filtrado de las entrevistas de los participantes

Se submuestreo previamente los archivos de audio .wav entre un rango mínimo de corte y rango máximo de corte, que [2.000.000,4.000.000], donde las muestras de audio se acercaron más entre sí al hacer el submuestreo [94], solucionando parcialmente el problema entre las entrevistas muy largas y las entrevistas muy cortas de los participantes.

Posteriormente se empleó un remuestreo [95], llevando las muestras de los audios de las entrevistas de los participantes a 5.000.000 de muestras de audio por participante, igualando las entrevistas y solucionando completamente el problema entre las entrevistas muy largas y las entrevistas muy cortas de los participantes.

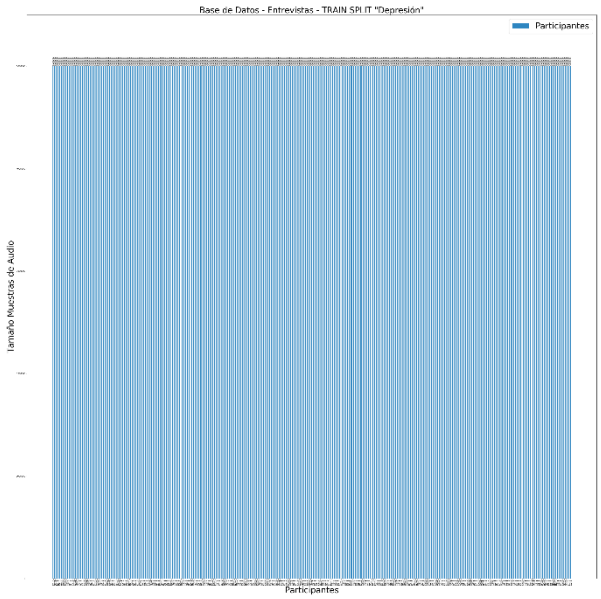
Se utilizó la librería científica SciPy[26] de python, importando la función signal (`from scipy import signal`), haciendo uso de Signal Processing Tools (Caja de Herramientas de Procesamiento de Señales) [96], más específicamente la función `resample`, para remuestrear las muestras utilizando el método de Fourier a lo largo de un eje dado, la función `resample` usa FFT[27] para volver a muestrear la señal de audio.

Dentro de los conjuntos de datos de train, dev, test, después de hacer el Remuestreo de las muestras de audio para los participantes de cada conjunto de datos se posee la siguiente información, tanto para los participantes depresivos, como para los participantes que no tienen depresión, como se observa a continuación:

- Train: El conjunto de datos de las Muestras de Audio de los Participantes de Train presenta para el conjunto de Train Depresivo, como para el conjunto de Train No Depresivo el mismo tamaño de muestras de audio para cada participante siendo esta 5.000.000, ver Gráfica 43 y Gráfica 44.

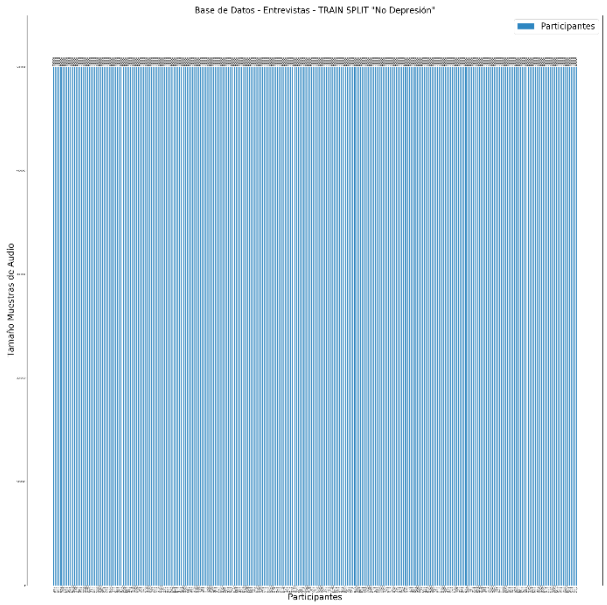
[26] SciPy (Scientific Python): Librería científica, libre y de código abierto para python, que se compone de herramientas y algoritmos matemáticos, una librería fundamental para la informática científica [111].

[27] FFT (Fast Fourier Transform): Es un algoritmo para el cálculo eficaz de la transformada Discreta de Fourier, que reduce el tiempo de cálculo de DFT significativamente, siendo un algoritmo computacional que divide la DFT en varias DFT más pequeñas, con el fin de descomponer señales, aplicar filtros, entre otras funciones [112].



Gráfica 43 - Muestras de Audio - Entrevistas - TRAIN "Depresión" Remuestreo

Fuente: Autoría propia, Edward Camilo Villota Taramuel

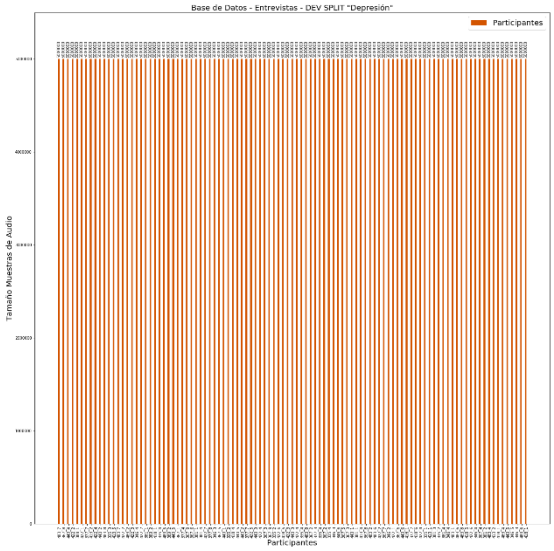


Gráfica 44 - Muestras de Audio - Entrevistas - TRAIN "No Depresión" Remuestreo

Fuente: Autoría propia, Edward Camilo Villota Taramuel

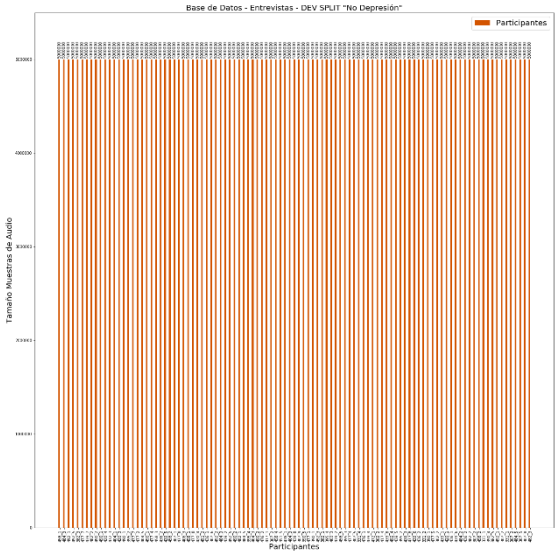
- Dev: El conjunto de datos de las Muestras de Audio de los Participantes de Dev presenta para el conjunto de Dev

Depresivo, como para el conjunto de Dev No Depresivo el mismo tamaño de muestras de audio para cada participante siendo esta 5.000.000, ver Gráfica 45 y Gráfica 46.



Gráfica 45 - Muestras de Audio - Entrevistas - DEV "Depresión" Remuestreo

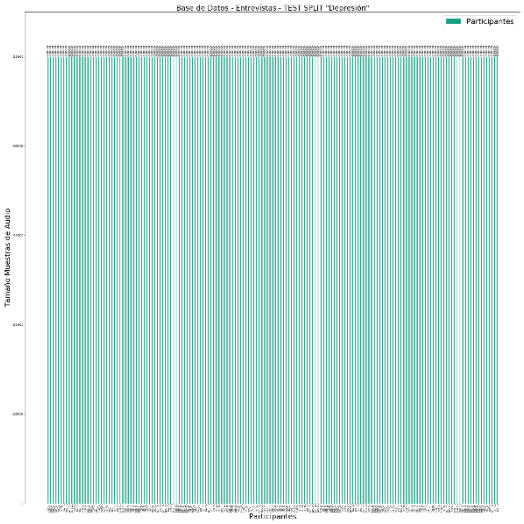
Fuente: Autoría propia, Edward Camilo Villota Taramuel



Gráfica 46 - Muestras de Audio - Entrevistas - DEV "No Depresión" Remuestreo

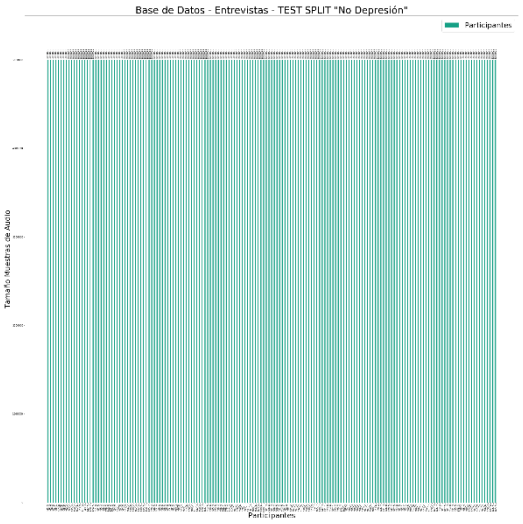
Fuente: Autoría propia, Edward Camilo Villota Taramuel

- Test: El conjunto de datos de las Muestras de Audio de los Participantes de Test presenta para el conjunto de Test Depresivo, como para el conjunto de Test No Depresivo el mismo tamaño de muestras de audio para cada participante siendo esta 5.000.000, ver Gráfica 47 y Gráfica 48.



Gráfica 47 - Muestras de Audio - Entrevistas - TEST "Depresión" Remuestreo

Fuente: Autoría propia, Edward Camilo Villota Taramuel



Gráfica 48 - Muestras de Audio - Entrevistas - TEST "No Depresión" Remuestreo

Fuente: Autoría propia, Edward Camilo Villota Taramuel

Para la realización de lo anterior, se empleó un código realizado en Python en un notebook de Jupyter Notebook llamado Graph_for_Split_Train_and_Test.ipynb.

8.3.2. Asignación del Conjunto de Datos de Entrenamiento y de Prueba

Para la selección del conjunto de entrenamiento y de prueba de la red convolucional, teniendo en cuenta el procesamiento y ajuste de la base de datos DAIC-WOZ realizada anteriormente, solucionando los problemas antes planteados, se utilizaron cuatro matrices para almacenar la información de las conversaciones de los participantes, además de las etiquetas binarias de depresión para Entrenamiento y Prueba, haciendo uso de la clase `np.memmap()`, dicha clase se utilizó para crear un mapeo de memoria en una matriz almacenada en un archivo binario en el disco, dado que dichas matrices poseían un tamaño demasiado grande para ser procesadas en memoria RAM.

El mapeo de memoria permite trabajar con matrices enormes casi como si fueran matrices normales. Python acepta tanto el manejo de matrices NumPy como entrada, al igual que matrices memmap. Sin embargo, se debe asegurar que la matriz memmap se utilice de manera eficiente. Es decir, la matriz nunca se carga como un todo (de lo contrario, desperdiciaría memoria del sistema y descartaría cualquier ventaja de la técnica), para ellos se debe acceder a los archivos mapeados en memoria utilizando pequeños segmentos de archivos grandes en el disco, sin leer todo el archivo en la memoria [97].

Se utilizó un DataFrame para almacenar la información de cada una las conversaciones .wav de las carpetas train_participant_1, train_participant_0, dev_participant_1, dev_participant_0, full_test_participant_1, full_test_participant_0, cargamos los archivos.wav haciendo uso de la librería Librosa, de dichos DataFrames por consiguiente se tomó los archivos que hacen alusión a train y test, seguidamente se unificaron en dos DataFrame, uno para train (depresivo, no depresivo) ver Tabla 25 y otro para test (depresivo, no depresivo) ver Tabla 26, teniendo una proporción de 50/50 de cada clase (depresivo, no depresivo)

Tabla 25 - DataFrame – Train (Depresivo, No Depresivo) para CNN

	participant	phq8_binary	sample	audio_sample
0	441_2	1	5000000	[-0.002627049, -0.0038023447, -0.004975671, -0...
1	372_4	1	5000000	[0.0016504877, 0.0018476367, 0.0020179672, 0.0...
...	...	...	...	...
640	468_4	0	5000000	[-0.017011574, -0.018964034, -0.020861235, -0....
641	315_3	0	5000000	[0.02360151, 0.025713377, 0.02751929, 0.028992...

Tabla 26 - DataFrame – Test (Depresivo, No Depresivo) para CNN

	participant	phq8_binary	sample	audio_sample
0	410_1	1	5000000	[0.003843443, 0.003009147, 0.0022143312, 0.001...
1	308_3	1	5000000	[0.008510126, 0.0024264602, -0.0033366322, -0....
...	...	...	...	...
336	466_5	0	5000000	[0.0066888495, 0.005923833, 0.005207789, 0.004...
337	465_5	0	5000000	[-4.7786827e-05, 4.145086e-05, 0.00011987491, ...

Dicha información se guardó en las matrices memmap X_train y X_test, respectivamente.

Las etiquetas para los Conjuntos de Datos de Train y Test se obtuvieron mediante Keras utilizando `from keras.utils.np_utils import to_categorical`, la cual tomó los valores de las columnas "phq8_binary", convirtiendo eso vales de vector de clase de (enteros) a una matriz de clase binaria [98], dichas información se guardó en las matrices memmap y_train y y_test, respectivamente.

Dando como resultado el siguiente conjunto de matrices:

- Entrenamiento:
  Se obtuvieron dos matrices, una de ellas es una matriz de 642 filas que hace alusión a los participantes y 5.000.000 de columnas que hace alusión a las

muestras de audio por participante, la otra matriz de 642 filas y 2 columnas que contienen la información de las etiquetas (depresivo, no depresivo), ilustradas en la Gráfica 49 y Gráfica 50.

```
memmap([[-0.00262705, -0.00380234, -0.00497567, ...,  0.0006899 ,
         -0.0003604 , -0.00147256],
        [ 0.00165049,  0.00184764,  0.00201797, ...,  0.00093435,
          0.00118894,  0.00142945],
        [ 0.01018338,  0.00929274,  0.00845859, ...,  0.01304765,
          0.01207675,  0.01111653],
        ...,
        [-0.00090339, -0.00105178, -0.00120368, ..., -0.00050144,
         -0.00062639, -0.0007609 ],
        [-0.01701157, -0.01896403, -0.02086123, ..., -0.01119484,
         -0.0130896 , -0.01504085],
        [ 0.02360151,  0.02571338,  0.02751929, ...,  0.01591799,
          0.01864278,  0.02122604]])
```

Gráfica 49 - Conjunto de Datos – Entrenamiento (642,5.000.000) CNN

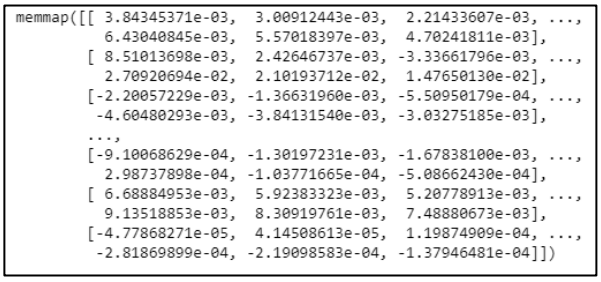
Fuente: Autoría propia, Edward Camilo Villota Taramuel

```
memmap([[0., 1.],
        [0., 1.],
        [0., 1.],
        ...,
        [1., 0.],
        [1., 0.],
        [1., 0.]], dtype=float32)
```

Gráfica 50 - Etiquetas para Conjunto de Datos – Entrenamiento (642,2) CNN

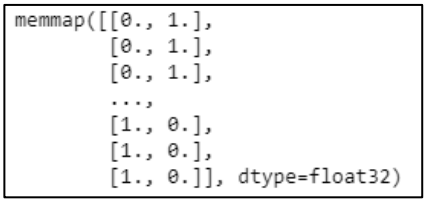
Fuente: Autoría propia, Edward Camilo Villota Taramuel

- Prueba:
  Se obtuvieron dos matrices, una de ellas es una matriz de 338 filas que hace alusión a los participantes y 5.000.000 de columnas que hace alusión a las muestras de audio por participante, la otra matriz de 338 filas y 2 columnas que contienen la información de las etiquetas (depresivo, no depresivo), ilustradas en la Gráfica 51 y Gráfica 52.



Gráfica 51 - Conjunto de Datos – Prueba (338,5.000.000) CNN

Fuente: Autoría propia, Edward Camilo Villota Taramuel



Gráfica 52 - Etiquetas para Conjunto de Datos – Prueba (338,2) CNN

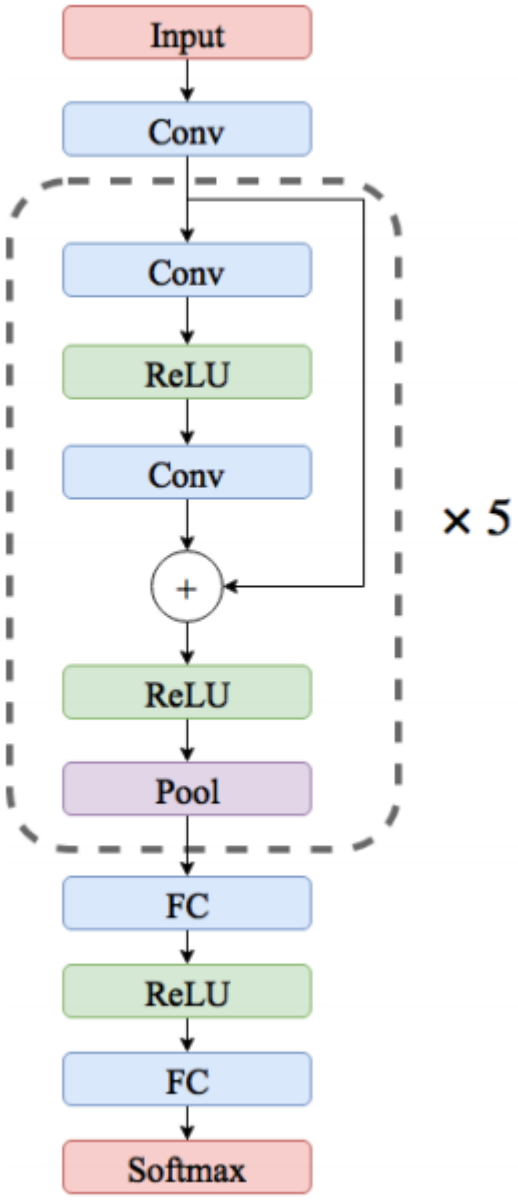
Fuente: Autoría propia, Edward Camilo Villota Taramuel

Para la realización de lo anterior, se empleó un código realizado en Python en un notebook de Jupyter Notebook llamado Create_Split_Train_and_Test.ipynb, dejando la información resultante de las matrices en archivos .dat, llamadas `X_train`, `y_train`, `X_test`, `y_test`, los cuales contienen toda la información para entrar en la CNN.

8.4. Implementar la arquitectura del modelo DepressionDetect

En el Aprendizaje Supervisado los datos para el entrenamiento incluyen la solución deseada, llamada etiquetas o labels [49]. El modelo aprende a base de unos datos etiquetados proporcionados previamente, donde el modelo toma como entrada unos datos y devuelve un resultado o predicción sobre esos datos adaptándose para dar una salida esperada de acuerdo con los datos de entrada.

La arquitectura de la Red Neuronal Convolucional que se usó está compuesta por varias capas que implementan la extracción de funciones o características y posteriormente realiza la clasificación de la depresión [64, p. 3], ver Gráfica 53.



Gráfica 53 - Arquitectura de la CNN *[64, p. 3]*

Se utilizan como entradas para la CNN las matrices anteriormente creadas con la información necesaria de las entrevistas de los participantes, la CNN comienza con una capa de entrada que se convierte en convolucionaria de 1-D con 32 filtros de tamaño 5, para crear 32 mapas de características.

Seguidos de cinco bloques residuales que contienen cada uno de ellos, una capa de convolución de 1-D con 32 filtros de tamaño 5 para crear 32

mapas de características seguidos de una función de activación ReLU y una segunda capa convolucional similar con filtros de 32 de tamaño 5, seguidos de una función de activación ReLU. A continuación los mapas de características experimentan una reducción de dimensionalidad con una capa de agrupación máxima (Max Pooling) de tamaño 5 con paso 2.

Por último, dos capas completamente conectadas (full-connected) o capas densas, acompañadas de una función de activación ReLU y una capa con una función softmax, que devuelve la probabilidad de que los datos ingresados esté en la clase depresiva o no depresiva.

8.5. Entrenar el modelo implementado con el conjunto de datos de entrenamiento

Se cargó en el sistema los archivos .dat (X_train, y_train, X_test, y_test) en los cuales esta almacenado la información relevante de las entrevistas de los audios de los participantes y las etiquetas (depresivo, no depresivo), los cuales constan de:

- X_train, y_train:
  Dos matrices, una de ellas es una matriz de 642 filas que hace alusión a los participantes y 5.000.000 de columnas que hace alusión a las muestras de audio por participante, la otra matriz de 642 filas y 2 columnas que contienen la información de las etiquetas (depresivo, no depresivo)

- X_test, y_test:
  Dos matrices, una de ellas es una matriz de 338 filas que hace alusión a los participantes y 5.000.000 de columnas que hace alusión a las muestras de audio por participante, y la otra matriz de 338 filas y 2 columnas que contienen la información de las etiquetas (depresivo, no depresivo)

El entrenamiento se realizó en CPU, teniendo limitaciones para procesar matrices de enorme tamaño (642 , 5.000.000) y (338, 5.000.000), donde el consumo de recursos en el ordenador y en su memoria RAM, superaban sus límites, haciendo muy difícil el entrenamiento del modelo, en consecuencia, se tomó conjuntos de prueba entre rangos de 100.000 mil muestras de audio, teniendo así varias instancias a lo largo de las muestra de audio de 5.000.000 millones, para cada participante, tomando los siguientes rangos:

- Conjuntos de entrenamiento y prueba para el rango de 0 a 100.000 muestras de audio:

Se tomo el rango de 0 a 100.000 muestras de audio, quedando los datos de entrenamiento y de prueba con matrices de (642,100.000) y (338,100.000) respectivamente

- Conjuntos de entrenamiento y prueba para el rango de 1.000.000 a 1.100.000 muestras de audio:
  Se tomo el rango de 1.000.000 a 1.100.000 muestras de audio, quedando los datos de entrenamiento y de prueba con matrices de (642,100.000) y (338,100.000) respectivamente

- Conjuntos de entrenamiento y prueba para el rango de 2.000.000 a 2.100.000 muestras de audio:
  Se tomo el rango de 2.000.000 a 2.100.000 muestras de audio, quedando los datos de entrenamiento y de prueba con matrices de (642,100.000) y (338,100.000) respectivamente

- Conjuntos de entrenamiento y prueba para el rango de 3.000.000 a 3.100.000 muestras de audio:
  Se tomo el rango de 3.000.000 a 3.100.000 muestras de audio, quedando los datos de entrenamiento y de prueba con matrices de (642,100.000) y (338,100.000) respectivamente

- Conjuntos de entrenamiento y prueba para el rango de 4.000.000 a 4.100.000 muestras de audio:
  Se tomo el rango de 4.000.000 a 4.100.000 muestras de audio, quedando los datos de entrenamiento y de prueba con matrices de (642,100.000) y (338,100.000) respectivamente

De esta manera haciendo posible el entrenamiento del modelo, con dichos rangos de muestras de audio, obteniendo 5 modelos con información diferente.

En esta etapa cada modelo se entrenó durante 50 épocas, con el rango de muestras de audio del conjunto de entrenamiento correspondiente, imprimiendo el número de la época en la que se encontraba el entrenamiento y la perdida y exactitud que se iba generando, como se observa a continuación:

```
# Train Model
history = model.fit(X_train, y_train, epochs=num_epochs, batch_size=batch_size, validation_data=(X_test, y_test))
```

```
Epoch 1/50
65/65 [==============================] - 569s 9s/step - loss: 0.5925 - accuracy: 0.6667 - val_loss: 0.9755 - val_accuracy: 0.5237
Epoch 2/50
65/65 [==============================] - 554s 9s/step - loss: 0.3546 - accuracy: 0.8474 - val_loss: 1.0374 - val_accuracy: 0.6065
Epoch 3/50
65/65 [==============================] - 554s 9s/step - loss: 0.2683 - accuracy: 0.9159 - val_loss: 1.2623 - val_accuracy: 0.5680
Epoch 4/50
65/65 [==============================] - 552s 8s/step - loss: 0.1172 - accuracy: 0.9548 - val_loss: 3.6431 - val_accuracy: 0.5325
Epoch 5/50
65/65 [==============================] - 552s 8s/step - loss: 0.2632 - accuracy: 0.9252 - val_loss: 2.6908 - val_accuracy: 0.5592
Epoch 6/50
65/65 [==============================] - 552s 8s/step - loss: 0.0650 - accuracy: 0.9875 - val_loss: 4.7726 - val_accuracy: 0.5385
Epoch 7/50
65/65 [==============================] - 552s 8s/step - loss: 0.0217 - accuracy: 0.9969 - val_loss: 5.9040 - val_accuracy: 0.5710
```

Gráfica 54 - Entrenamiento del Modelo - CNN

Fuente: Autoría propia, Edward Camilo Villota Taramuel

8.6. Comprobar el modelo implementado con el conjunto de datos de prueba disponibles en la base de datos

En la CNN después del entrenamiento del modelo, se procedió, a evaluar el modelo, con el conjunto de datos de entrenamiento de cada rango de muestras de audio correspondiente.

Seguidamente se evaluó el rendimiento de cada modelo, para verificar la exactitud del modelo y se creó una matriz de confusión para cada modelo para verificar la precisión de dicho modelo.

Los modelos de detección del estado de la depresión creados exhiben un accuracy de 0.54, utilizando las características prosódicas del habla de una persona. El modelo desarrollado aún no se encuentra en un estado predictivo para uso recomendado, aunque estos resultados sugieren fuertemente una nueva y prometedora dirección para la detección de la depresión.

Debido al tamaño de los datos no fue posible correr el modelo con todos estos por lo que se limitó el tamaño de muestra a 100.000. Sin embargo, como un elemento de verificación se corrió el modelo con muestras de este tamaño en ubicaciones diferentes de los datos, específicamente en los rangos 0 a 100.000, 1.000.000 a 1.100.000, 2.000.000 a 2.100.000, 3.000.000 a 3.100.000, 4.000.000 a 4.100.000, en base a los cuales se obtuvieron los resultados, que se muestran en la Tabla 27, Tabla 28, Tabla 29, Tabla 30, Tabla 31, Tabla 32, Tabla 33, Tabla 34, Tabla 35, Tabla 36.

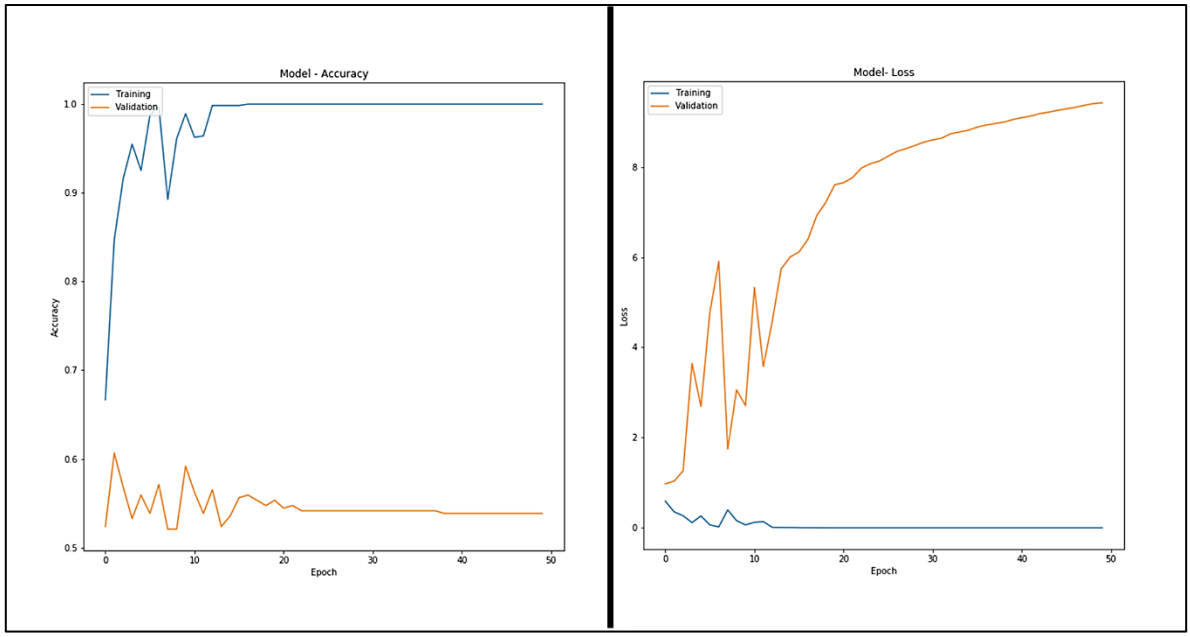
- Modelo CNN - rango de 0 a 100.000 muestras de audio

Tabla 27 - Resultados del Modelo – Matriz de Confusión - rango de 0 a 100.000 muestras de audio

Tabla 1: Predicciones de conjuntos de prueba para el rango de 0 a 100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	125 (VP)	44 (FP)
Pronosticada: No	112 (FP)	57 (TN)

Tabla 28 - Resultados del Modelo - rango de 0 a 100.000 muestras de audio

Tabla 2: Predicciones de conjuntos de prueba para el rango de 0 a 100.000 muestras de audio			
F1 score 0.62	precision 0.53	recall 0.74	accuracy 0.54



Gráfica 55 - Épocas - exactitud (accuracy) y perdida (loss) del Modelo - rango de 0 a 100.000 muestras de audio - CNN

Fuente: Autoría propia, Edward Camilo Villota Taramuel

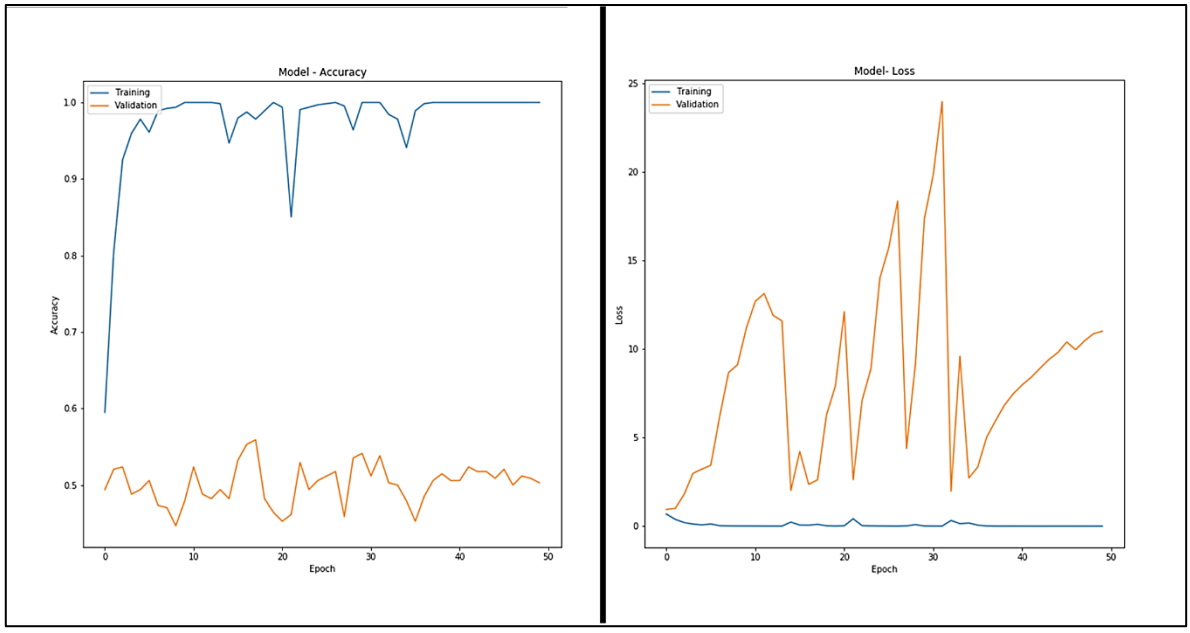
- Modelo CNN - rango de 1.000.000 a 1.100.000 muestras de audio

Tabla 29 - Resultados del Modelo – Matriz de Confusión - rango de 1.000.000 a 1.100.000 muestras de audio

Tabla 3: Predicciones de conjuntos de prueba para el rango de 1.000.000 a 1.100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	130 (VP)	39 (FN)
Pronosticada: No	129 (FP)	40 (VN)

Tabla 30 - Resultados del Modelo - rango de 1.000.000 a 1.100.000 muestras de audio

Tabla 4: Predicciones de conjuntos de prueba para el rango de 1.000.000 a 1.100.000 muestras de audio			
F1 score 0.61	precision 0.50	recall 0.77	accuracy 0.50



Gráfica 56 - Épocas - exactitud (accuracy) y perdida (loss) del Modelo - rango de 1.000.000 a 1.100.000 muestras de audio – CNN

Fuente: Autoría propia, Edward Camilo Villota Taramuel

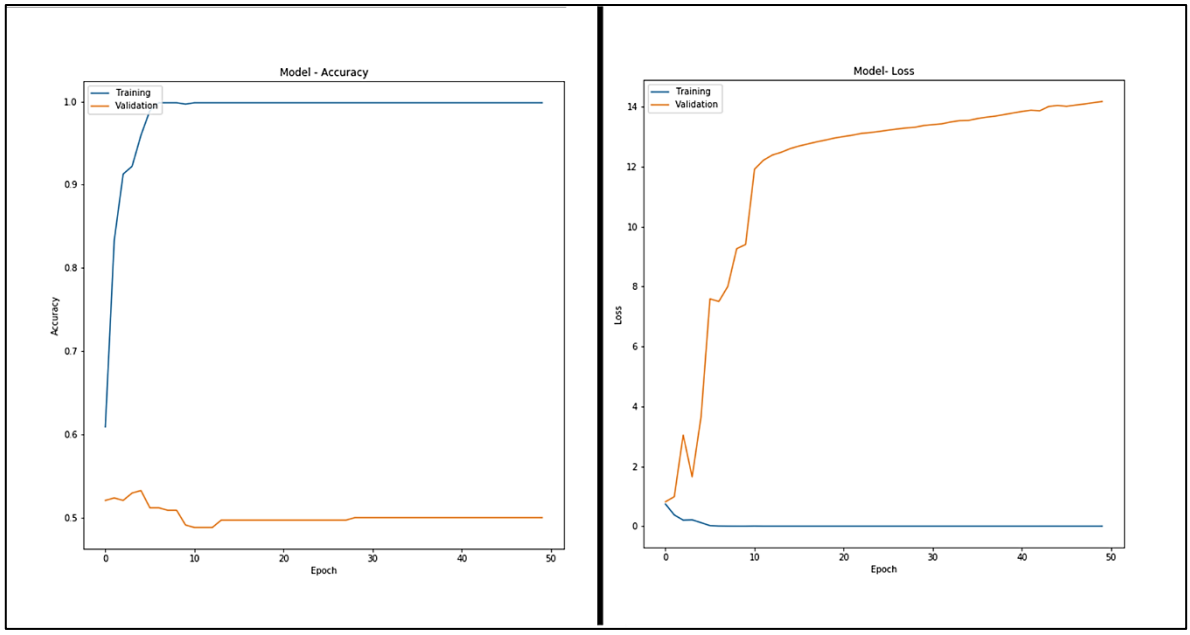
- Modelo CNN - rango de 2.000.000 a 2.100.000 muestras de audio

Tabla 31 - Resultados del Modelo – Matriz de Confusión - rango de 2.000.000 a 2.100.000 muestras de audio

Tabla 5: Predicciones de conjuntos de prueba para el rango de 2.000.000 a 2.100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	142 (VP)	27 (FN)
Pronosticada: No	142 (FP)	27 (VN)

Tabla 32 - Resultados del Modelo - rango de 2.000.000 a 2.100.000 muestras de audio

Tabla 6: Predicciones de conjuntos de prueba para el rango de 2.000.000 a 2.100.000 muestras de audio			
F1 score 0.63	precision 0.50	recall 0.84	accuracy 0.50



Gráfica 57 - Épocas - exactitud (accuracy) y perdida (loss) del Modelo - rango de 2.000.000 a 2.100.000 muestras de audio – CNN

Fuente: Autoría propia, Edward Camilo Villota Taramuel

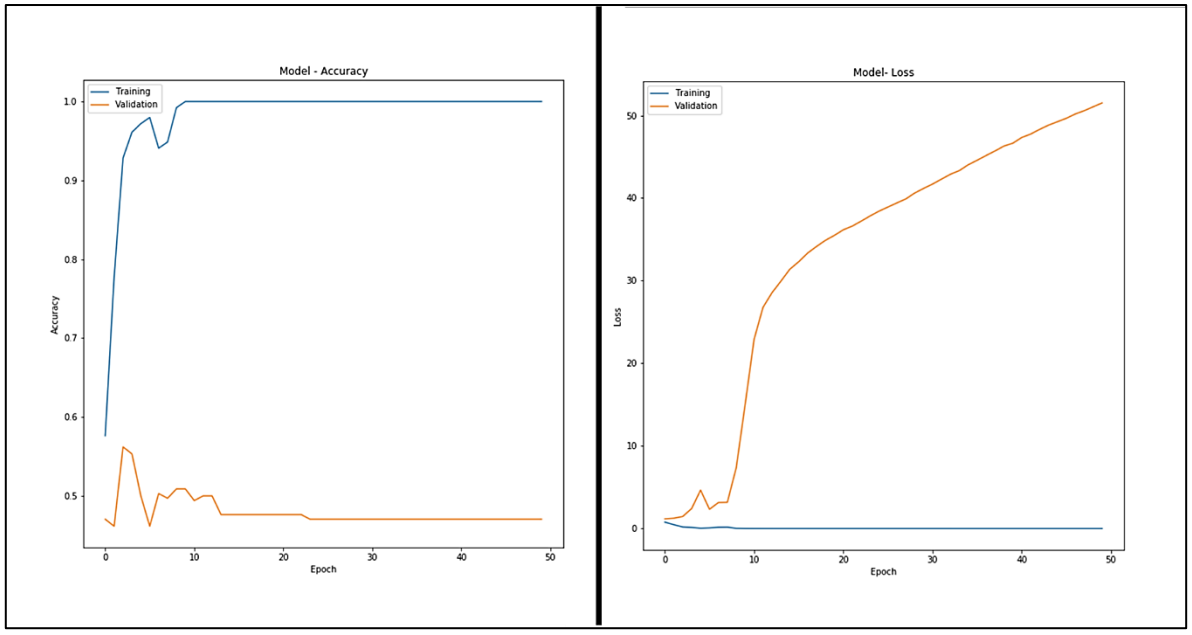
- Modelo CNN - rango de 3.000.000 a 3.100.000 muestras de audio

Tabla 33 - Resultados del Modelo – Matriz de Confusión - rango de 3.000.000 a 3.100.000 muestras de audio

Tabla 7: Predicciones de conjuntos de prueba para el rango de 3.000.000 a 3.100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	132 (VP)	37 (FN)
Pronosticada: No	142 (FP)	27 (VN)

Tabla 34 - Resultados del Modelo - rango de 3.000.000 a 3.100.000 muestras de audio

Tabla 8: Predicciones de conjuntos de prueba para el rango de 3.000.000 a 3.100.000 muestras de audio			
F1 score 0.60	precision 0.48	recall 0.78	accuracy 0.47



Gráfica 58 - Épocas - exactitud (accuracy) y perdida (loss) del Modelo - rango de 3.000.000 a 3.100.000 muestras de audio – CNN

Fuente: Autoría propia, Edward Camilo Villota Taramuel

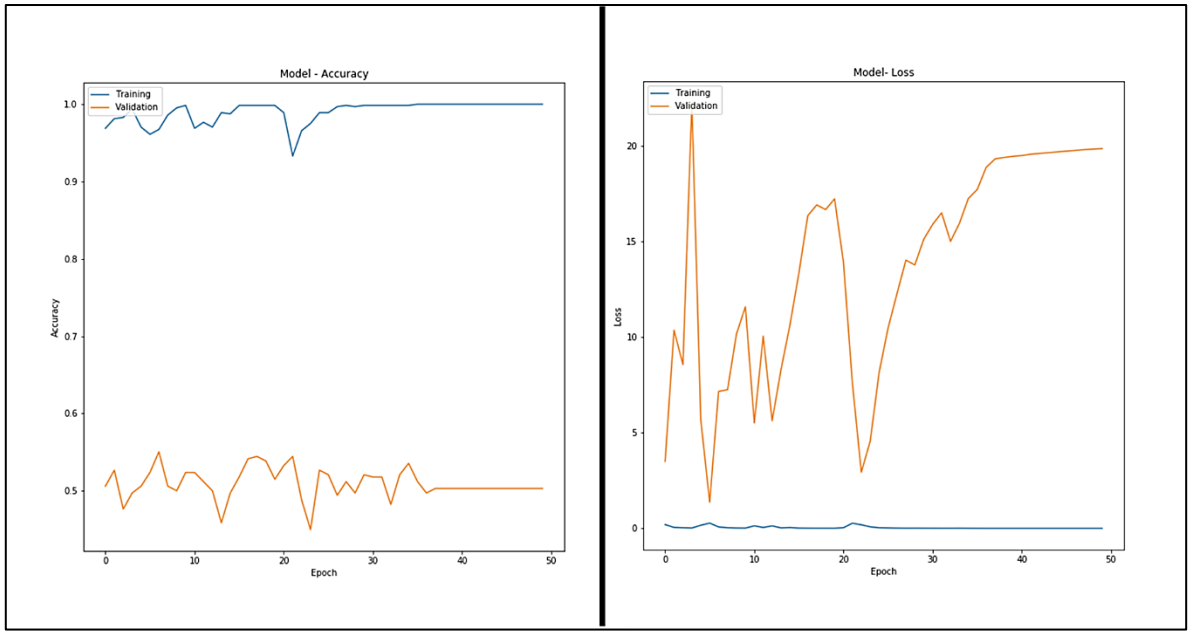
- Modelo CNN - rango de 4.000.000 a 4.100.000 muestras de audio

Tabla 35 - Resultados del Modelo – Matriz de Confusión - rango de 4.000.000 a 4.100.000 muestras de audio

Tabla 9: Predicciones de conjuntos de prueba para el rango de 4.000.000 a 4.100.000 muestras de audio		
Confusion Matrix	Actual: Si	Actual: No
Pronosticada: Si	115 (VP)	54 (FN)
Pronosticada: No	114 (FP)	55 (VN)

Tabla 36 - Resultados del Modelo - rango de 4.000.000 a 4.100.000 muestras de audio

Tabla 10: Predicciones de conjuntos de prueba para el rango de 4.000.000 a 4.100.000 muestras de audio			
F1 score 0.58	precision 0.50	recall 0.68	accuracy 0.50



Gráfica 59 - Épocas - exactitud (accuracy) y perdida (loss) del Modelo - rango de 4.000.000 a 4.100.000 muestras de audio - CNN

Fuente: Autoría propia, Edward Camilo Villota Taramuel

9. RESULTADOS

Luego de realizados el entrenamiento y comprobación de los modelos correspondientes, teniendo en cuenta que se realizaron con diferentes rangos de muestras de audio de las entrevistas de los participantes de la Base de Datos DAIC-WOZ, se obtuvieron los siguientes resultados, ver Tabla 37.

Tabla 37 - Resultados

Conjunto de Datos	Exactitud del Modelo (Accuracy)
Modelo 0 - Rango de 0 a 100.000 muestras de audio	53.85%
Modelo 1 - Rango de 1.000.000 a 1.100.000 muestras de audio	50.30%
Modelo 2 - Rango de 2.000.000 a 2.100.000 muestras de audio	50.00%
Modelo 3 - Rango de 3.000.000 a 3.100.000 muestras de audio	47.04%
Modelo 4 - Rango de 4.000.000 a 4.100.000 muestras de audio	50.30%

Al analizar los resultados con las muestras de audio en cierto rango, es posible observar que el modelo de detección del estado de la depresión creado exhibe un accuracy de 0.54, siendo eficaz en cierta medida, no obstante mejorable considerablemente, ya que se obtuvo 54 % de exactitud en el reconocimiento de la depresión.

Los resultados obtenidos se encuentran por debajo de los resultados del modelo DepressionDetect, ya que este posee un accuracy de 0.58, no obstante al presentar dificultad debido a las limitaciones computacionales durante la fase de entrenamiento y de prueba de la red neuronal convolucional, al ejecutarla sobre CPU y al tener que limitar en número de muestras de audio a entrenar, adaptándolas a los recursos de procesamiento que contaba el PC, se vio reducido el nivel de accuracy del modelo.

Siendo, la ejecución de las fases de entrenamiento y de prueba de la CNN sobre GPU más óptima para dicho proceso, donde el flujo de trabajo con GPU es varias veces más fluido que solamente con CPU, utilizando miles de núcleos más pequeños y eficientes para una arquitectura paralela masiva dirigida a manejar múltiples funciones al mismo tiempo [99].

Ciertamente el modelo DepressionDetect desarrollado para el diagnóstico automático de la depresión mediante voz aún no se encuentra en un estado predictivo para uso recomendado, pero estos resultados sugieren fuertemente una nueva y prometedora dirección para la detección de la depresión, haciendo uso de las características prosódicas del habla de una persona.

CONCLUSIONES

- En este proyecto se validó un modelo para la clasificación de la depresión mediante la voz entrenando una red neuronal convolucional profunda. Según los resultados, el modelo aun no es capaz de ser usado como estado predictivo para la depresión por su bajo puntaje de exactitud al momento de detectar la enfermedad, a pesar de esto, se demostró que las características prosódicas del habla de una persona pueden ser predictores prometedores de la depresión.

- Este tipo de modelos pueden ser de gran utilidad como una propuesta para reducir la carga de los médicos para diagnosticar la depresión, respaldar los diagnósticos de los profesionales médicos, contribuyendo desde el punto de vista social al campo de la salud al mejorar las condiciones de diagnóstico del paciente proporcionando una evaluación más objetiva y un diagnóstico rápido a través de redes neuronales convolucionales, pudiendo convertirse en un aporte a la Salud y al análisis de enfermedades mentales.

RECOMENDACIONES

Dentro de un proyecto tan ambicioso como lo fue éste, siempre se desea que haya una mejora continua, en pro al beneficio y crecimiento de la detección de la depresión en términos de detección automática mediante redes neuronales convolucionales. A pesar de que se lograron los objetivos planteados en el proyecto, aún queda mucho trabajo por hacer, así como modelos que pueden ser explorados.

Por lo tanto, se sugiere para futuros trabajos:

- Intentar modificar la arquitectura de la red neuronal convolucional o hacer uso de algunas arquitecturas más robustas

- Variar el tamaño del conjunto de datos que se emplean para entrenamiento y prueba de la CNN.

- Realizar la ejecución de las fases de entrenamiento y de prueba de la CNN sobre GPU aprovechando el flujo de trabajo más fluido y haciendo uso de una arquitectura paralela

Con el fin de aumentar la exactitud del modelo

BIBLIOGRAFÍA

[1] L. Toro, «Anaconda Distribution: La Suite más completa para la Ciencia de datos con Python,» 14 09 2017. [En línea]. Available: https://blog.desdelinux.net/ciencia-de-datos-con-python/. [Último acceso: 02 08 2020].

[2] OMS, «Atención primaria de salud,» [En línea]. Available: https://www.who.int/topics/primary_health_care/es/. [Último acceso: 14 06 2019].

[3] F. J. Rogel-Ortiz, «Autismo,» 22 07 2004. [En línea]. Available: http://www.scielo.org.mx/pdf/gmm/v141n2/v141n2a9.pdf. [Último acceso: 16 06 2019].

[4] K. Kiefer, «DepressionDetect: A Machine Learning Approach for Audio based Depression Classification,» 20 06 2017. [En línea]. Available: https://github.com/kykiefer/depression-detect/blob/master/depression-detect-report.pdf. [Último acceso: 18 06 2019].

[5] G. Degottex, J. Kane, T. Drugman, T. Raitio y S. Scherer, «COVAREP – A COLLABORATIVE VOICE ANALYSIS REPOSITORY FOR SPEECH TECHNOLOGIES,» 09 05 2014. [En línea]. Available: https://ieeexplore.ieee.org/document/6853739. [Último acceso: 20 02 2020].

[6] L. Parra, «Qué es un csv, cómo se hace y para qué sirve,» 14 01 2015. [En línea]. Available: https://lolap.wordpress.com/2015/01/14/que-es-un-csv-como-se-hace-y-para-que-sirve/. [Último acceso: 02 08 2020].

[7] USC - University of Southern California, «DAIC-WOZ Database,» [En línea]. Available: http://dcapswoz.ict.usc.edu/. [Último acceso: 18 06 2019].

[8] T. Balagueró, «¿Qué son los datasets y los dataframes en el Big Data?,» 13 11 2018. [En línea]. Available: https://www.deustoformacion.com/blog/programacion-diseno-web/que-son-datasets-dataframes-big-data#:~:text=%C2%BFQu%C3%A9%20es%20un%20dataset%3F,colecci%C3%B3n%20de%20datos%20habitualmente%20tabulada.. [Último acceso: 02 08 2020].

[9] Organización Panamericana de la Salud, «Demencia: una prioridad de salud pública,» 2013. [En línea]. Available: https://apps.who.int/iris/bitstream/handle/10665/98377/9789275318256_spa.pdf;jsessionid=34270FEB1EB6946F36C30FF8E10D9D5B?sequence=1. [Último acceso: 18 06 2019].

[10] Organización Panamericana de la Salud, «Depresión y otros trastornos mentales comunes. Estimaciones sanitarias mundiales,» 2017. [En línea]. Available: https://iris.paho.org/bitstream/handle/10665.2/34006/PAHONMH17005-spa.pdf?sequence=1&isAllowed=y. [Último acceso: 18 06 2019].

[11] K. Xiaoyan y L. Jing , Discapacidad intelectual, Ginebra, 2017.

[12] OMS, «Discapacidad y salud,» 16 01 2018. [En línea]. Available: https://www.who.int/es/news-room/fact-sheets/detail/disability-and-health#:~:text=La%20Clasificaci%C3%B3n%20Internacional%20del%20Funcionamiento,y%20restricciones%20a%20la%20participaci%C3%B3n.. [Último acceso: 02 07 2019].

[13] OMS, «Informe sobre la salud en mundo 2001. Salud mental: nuevos conocimientos, nuevas esperanzas,» Ginebra, 2001.

[14] W. Jiménez Jiménez, «Salud mental en el posconflicto colombiano,» 28 05 2009. [En línea]. Available: http://www.scielo.org.co/pdf/crim/v51n1/v51n1a07.pdf. [Último acceso: 02 07 2019].

[15] musiki.org, «Frecuencia de muestreo,» 25 07 2019. [En línea]. Available: http://musiki.org.ar/Frecuencia_de_muestreo. [Último acceso: 02 08 2020].

[16] P. Compañ Rosique, M. Cazorla Quevedo, F. Escolano Ruiz y R. Rizo Aldeguer, «Fundamentos de inteligencia artificial», Alicante , España: Publicaciones de la Universidad de Alicante, 1999, p. 2.

[17] M. Gwydion, «Primeros pasos con Jupyter Notebook,» 18 01 2018. [En línea]. Available: https://www.adictosaltrabajo.com/2018/01/18/primeros-pasos-con-jupyter-notebook/. [Último acceso: 02 08 2020].

[18] librosa.org, «Librosa,» 2020. [En línea]. Available: https://librosa.org/doc/latest/index.html. [Último acceso: 02 08 2020].

[1 O. Llauradó, «La escala de Likert: qué es y cómo utilizarla,» 12 12 2014. [En 9] línea]. Available: https://www.netquest.com/blog/es/la-escala-de-likert-que-es-y-como-utilizarla. [Último acceso: 22 11 2019].

[2 C. Zapata y N. Carmona, «EL EXPERIMENTO MAGO DE OZ Y SUS 0] APLICACIONES: UNA MIRADA RETROSPECTIVA,» 15 10 2006 . [En línea]. Available: https://www.researchgate.net/publication/262507524_WIZARD-OF-OZ_EXPERIMENT_AND_ITS_APPLICATIONS_AN_OVERVIEW. [Último acceso: 18 06 2019].

[2 learnpython.org, «Pandas Basics,» 2019. [En línea]. Available: 1] https://www.learnpython.org/es/Pandas%20Basics. [Último acceso: 02 08 2020].

[2 K. Kroenke, T. W. Strine, R. L. Spitzer, J. W. Williams y J. T. Be, «The PHQ-8 as 2] a measure of current depression in the general population,» 27 08 2008. [En línea]. Available: https://indiana.pure.elsevier.com/en/publications/the-phq-8-as-a-measure-of-current-depression-in-the-general-popul. [Último acceso: 18 06 2019].

[2 A. Downey, J. Elkner y C. Meyers, «Aprenda a Pensar Como un Programador 3] con Python,» 04 2002. [En línea]. Available: https://argentinaenpython.com/quiero-aprender-python/aprenda-a-pensar-como-un-programador-con-python.pdf. [Último acceso: 02 08 2020].

[2 OMS, «Promocón de la salud mental: conceptos, evidencia emergente, 4] práctica,» Ginebra, 2004.

[2 tensorflow.org, «Por qué TensorFlow,» 2020. [En línea]. Available: 5] https://www.tensorflow.org/. [Último acceso: 02 08 2020].

[2 J. E. Dimsdale, «Trastorno de síntomas somáticos,» MANUAL MSD - Versión 6] para profesionales, 08 2019. [En línea]. Available: https://www.msdmanuals.com/es/professional/trastornos-psiqui%C3%A1tricos/trastornos-de-s%C3%ADntomas-som%C3%A1ticos-y-relacionados/trastorno-de-s%C3%ADntomas-som%C3%A1ticos. [Último acceso: 02 08 2020].

[2 masadelante.com, «¿Qué es WAV? - Definición de WAV,» 2019. [En línea]. 7] Available: https://www.masadelante.com/faqs/wav. [Último acceso: 02 08 2020].

[2 J. C. Mundt, P. J. Snyder, M. S. Cannizzarro, K. Chappie y D. S. Geralts, «Voice 8] acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology,» 04 04 2006. [En línea]. Available:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3022333/. [Último acceso: 05 07 2019].

[29] L. Jinming, F. Xiaoyan, S. Zhuhong y S. Yuanyuan , «Improvement on Speech Depression Recognition Based on Deep Networks,» 24 01 2019. [En línea]. Available: https://ieeexplore.ieee.org/document/8623055. [Último acceso: 14 06 2019].

[30] OMS, «Depresión,» 22 03 2018. [En línea]. Available: https://www.who.int/es/news-room/fact-sheets/detail/depression. [Último acceso: 14 06 2019].

[31] D.J.France, R.G.Shiavi, S.Silverman, D.M.Wilkes y M.Silverman, «Acoustical properties of speech as indicators of depression and suicidal risk,» 07 2000. [En línea]. Available: https://ieeexplore.ieee.org/document/846676. [Último acceso: 14 07 2019].

[32] S.Scherer, G.Stratou, J.Gratch y L.-P.Morency, «Investigating voice quality as a speaker-independent indicator of depression and ptsd,» 2013. [En línea]. Available: https://ict.usc.edu/pubs/Investigating%20Voice%20Quality%20as%20a%20Speaker-Independent%20Indicator%20of%20Depression%20and%20PTSD.pdf. [Último acceso: 14 07 2019].

[33] A. L. Rosset, M. A. García, E. Destéfanis, S. Cerruti y M. Moyano, «Deep Neural Network para Análisis Acústico,» 04 2018. [En línea]. Available: http://sedici.unlp.edu.ar/handle/10915/67072. [Último acceso: 14 07 2019].

[34] G. Hinton, L. Deng, D. Yu, G. Ddahl, Abdel-rahman, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath y B. Kingsbury, «Deep Neural Networks for Acoustic Modeling in Speech Recognition,» 27 04 2012. [En línea]. Available: https://static.googleusercontent.com/media/research.google.com/es//pubs/archive/38131.pdf. [Último acceso: 14 07 2019].

[35] J. M. Calvo Gómez y L. E. Jaramillo Gonzáles, «Detección del trastorno depresivo mayor en atención primaria. Una revisión,» 11 02 2015. [En línea]. Available: http://www.scielo.org.co/pdf/rfmun/v63n3/v63n3a15.pdf. [Último acceso: 22 07 2019].

[36] W. J. Stanley , Historia de la Melancolía y la Depresión, London: Ediciones Turner, S. A., 1986.

[37] A. Cia Puyuelo, «ANSIEDAD Y DEPRESIÓN,» 17 06 2013. [En línea]. Available: https://repositori.udl.cat/bitstream/handle/10459.1/46960/aciap.pdf?sequence=1&isAllowed=y. [Último acceso: 28 07 2019].

[38] Ministerio de Salud y Protección Social - Colciencias, «Guía de práctica clínica. Detección temprana y diagnóstico de depresión (episodio depresivo y trastorno depresivo recurrente) en adultos - 2013 Guía No. 22,» 04 2013. [En línea]. Available: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/INEC/IETS/GPC_Ptes_Depre%20(1).pdf. [Último acceso: 16 08 2019].

[39] Universidad Internacional de Valencia, «Depresión endógena: síntomas y causas,» 21 03 2018. [En línea]. Available: https://www.universidadviu.com/depresion-endogena-sintomas-causas/. [Último acceso: 17 08 2019].

[40] F. Vázquez, R. Muñoz y E. Becoña, «DEPRESIÓN: DIAGNÓSTICO, MODELOS TEÓRICOSY TRATAMIENTO A FINALES DEL SIGLO XX,» 2000. [En línea]. Available: https://www.academia.edu/37742068/DEPRESI%C3%93N_DIAGN%C3%93STICO_MODELOS_TE%C3%93RICOS_Y_TRATAMIENTO_A_FINALES_DEL_SIGLO_XX. [Último acceso: 17 08 2019].

[41] M. S. Batista Capelo, «CARACTERIZACIÓN DEL PERFIL DE LENGUAJE EN EL TRASTORNO DEPRESIVO Y EN EL DOLOR CRÓNICO,» 06 2019. [En línea]. Available: https://scholar.googleusercontent.com/scholar?q=cache:ol5ax0SC-b4J:scholar.google.com/&hl=es&as_sdt=0,5&scioq=CARACTERIZACI%C3%93N+DEL+PERFIL+DE+LENGUAJE+EN+EL+TRASTORNO+DEPRESIVO+Y+EN+EL+DOLOR+CR%C3%93NICO. [Último acceso: 12 11 2019].

[42] F. Escolano Ruiz, M. A. Cazorla , M. Galipienso, O. Colomina Pardo y M. M. Lozano Ortega, Inteligencia artificial: modelos, técnicas y áreas de aplicación, Paraninfo, 2003.

[43] L. Qiuhua, L. Xuejun y C. Lawrence, «Semi-Supervised Life-Long Learning with Application to Sensing,» 14 12 2007. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/4497950. [Último acceso: 16 08 2019].

[44] S. Dhankhad, E. Mohammed y B. Far, «Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study,» 06 07 2018. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/8424696. [Último acceso: 17 08 2019].

[4 X. Zhu y G. A., Introduction to Semi-Supervised Learning, Morgan & Claypool, 5] 2009.

[4 D. Happiness, Z. Yimin, D. Kumar y W. Qingtian, «Unsupervised Learning Based 6] On Artificial Neural Network: A Review,» 17 01 2019. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/8612259. [Último acceso: 17 08 2019].

[4 Q. Li, J. Zhao y X. Zhu, «An Unsupervised Learning Algorithm for Intelligent 7] Image Analysis,» 05 12 2007. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/4150129. [Último acceso: 17 08 2019].

[4 J. J. Bagnato, «Aprende Machine Learning,» 27 08 2017. [En línea]. Available: 8] https://www.aprendemachinelearning.com/que-es-machine-learning/. [Último acceso: 17 08 2019].

[4 J. J. Bagnato, «Aplicaciones del Machine Learning,» 05 09 2017. [En línea]. 9] Available: https://www.aprendemachinelearning.com/aplicaciones-del-machine-learning/#supervisado. [Último acceso: 17 08 2019].

[5 D. Erroz Arroyo, «Visualizando neuronas en Redes Neuronales 0] Convolucionales,» 10 06 2019. [En línea]. Available: https://academica-e.unavarra.es/xmlui/bitstream/handle/2454/33694/memoria_TFG.pdf?sequence=1&isAllowed=y. [Último acceso: 17 08 2019].

[5 R. Lopez Briega, «Introducción al Deep Learning,» 13 06 2017. [En línea]. 1] Available: https://relopezbriega.github.io/blog/2017/06/13/introduccion-al-deep-learning/. [Último acceso: 17 08 2019].

[5 A. Mosquera Mosquera y J. Martinez Rendon, «Reconocimiento Optico de 2] Caracteres en Placas Vehiculares haciendo uso de Redes Neuronales Convolucionales,» 11 2018. [En línea]. Available: http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/10027/T005.12%20M912.pdf?sequence=1&isAllowed=y. [Último acceso: 17 08 2019].

[5 F. J. Núñez Sánchez , «Diseño de un sistema de reconocimiento automático de 3] matrículas de vehículos mediante una red neuronal convolucional,» 01 06 2016. [En línea]. Available: http://openaccess.uoc.edu/webapps/o2/bitstream/10609/52222/7/fnunezsTFM0616memoria.pdf. [Último acceso: 18 08 2019].

[5 M. S. B. y Z. C. C., «Breast Cancer prediction based on Backpropagation 4] Algorithm,» 28 01 11. [En línea]. Available: https://ieeexplore-ieee-

org.ezproxy.utp.edu.co/document/5703994/authors#authors. [Último acceso: 19 10 01].

[5 5] H. Yanagisawa, T. Yamashita y H. Watanabe, «A study on object detection method from manga images using CNN,» 31 05 2018. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/8369633. [Último acceso: 18 08 2019].

[5 6] Z. Y., X. Xiao y X. Yang, «Real-Time Object Detection for 360-Degree Panoramic Image Using CNN,» 21 10 2017. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/8719156. [Último acceso: 18 08 2019].

[5 7] K. Murata, M. Mito, D. Eguchi, Y. Mori y M. Toyonaga, «A Single Filter CNN Performance for Basic Shape Classification,» 01 11 2019. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/8517219. [Último acceso: 18 08 2019].

[5 8] U. Amin, S. M. A., M. Bilal y R. M. M., «Classification of Arrhythmia by Using Deep Learning with 2-D ECG Spectral Image Representation,» 25 05 2020. [En línea]. Available: https://www.mdpi.com/2072-4292/12/10/1685/pdf. [Último acceso: 10 06 2020].

[5 9] G. DC, «Arrhythmia on ECG Classification using CNN,» 2019. [En línea]. Available: https://www.kaggle.com/gregoiredc/arrhythmia-on-ecg-classification-using-cnn/notebook. [Último acceso: 15 01 2020].

[6 0] W. Rawat y Z. Wang, «Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review.,» 29 09 2017. [En línea]. Available: https://www.ncbi.nlm.nih.gov/pubmed/28599112. [Último acceso: 18 08 2019].

[6 1] G. A. Martinez y G. Aguilar, «Reconocimiento de voz basado en MFCC, SBC y Espectrogramas,» 10 12 2013. [En línea]. Available: http://www.redalyc.org/pdf/5055/505554816003.pdf. [Último acceso: 18 08 2019].

[6 2] J. Miranda Orozco, «DESCRIPCIÓN DE LOS ELEMENTOS PROSÓDICOS DEL HABLA DE LA LENGUA DE SEÑAS VENEZOLANA,» 05 2016. [En línea]. Available: https://cultura-sorda.org/wp-content/uploads/2016/06/Tesis-Miranda-Interpretacion-LSV-2016.pdf. [Último acceso: 18 08 2019].

[6 3] D. Conde Ortiz, «Inteligencia Artificial con TensorFlow para predicción de comportamientos,» 2018. [En línea]. Available:

https://pdfs.semanticscholar.org/bf01/10c9822592997a8ecef35bb3963e86276c54.pdf. [Último acceso: 10 08 2020].

[64] M. Kachuee, S. Fazeli y M. Majid Sarrafzadeh, «ECG Heartbeat Classification: A Deep Transferable Representation,» 12 07 2018. [En línea]. Available: https://arxiv.org/pdf/1805.00794.pdf. [Último acceso: 06 08 2020].

[65] CONGRESO DE COLOMBIA, «Ley 1616 de 2013 - Salud Mental,» 13 01 2013. [En línea]. Available: https://www.asivamosensalud.org/politicas-publicas/normatividad-leyes/salud-publica/ley-1616-de-2013-salud-mental. [Último acceso: 20 08 2019].

[66] constitucioncolombia, «Artículo 49,» [En línea]. Available: http://www.constitucioncolombia.com/titulo-2/capitulo-2/articulo-49. [Último acceso: 20 08 2019].

[67] «LEY 1122 - 09/01/2007,» [En línea]. Available: https://www.funlam.edu.co/uploads/facultadpsicologia/790081.pdf. [Último acceso: 20 08 2019].

[68] «LEY 100 DE 1993,» 23 12 1993. [En línea]. Available: http://www.secretariasenado.gov.co/senado/basedoc/ley_0100_1993.html. [Último acceso: 20 08 2019].

[69] OCDE, «Instrumentos Legales de la OCDE,» [En línea]. Available: https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449. [Último acceso: 20 08 2019].

[70] R. Belmaker y A. G., «Major Depressive Disorder,» 03 01 2008. [En línea]. Available: https://www.nejm.org/doi/full/10.1056/nejmra073096. [Último acceso: 20 08 2019].

[71] H. M. Y. G. F. Z. A. Jan, «Artificial Intelligent System for Automatic Depression Level Analysis Through Visual and Vocal Expressions,» 31 07 2017. [En línea]. Available: https://ieeexplore.ieee.org/abstract/document/7997822. [Último acceso: 20 08 2019].

[72] E. G. a. A. H. W. C. de Melo, «Combining Global and Local Convolutional 3D Networks for Detecting Depression from Facial Expressions,» 11 07 2019. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/8756568. [Último acceso: 20 08 2019].

[73] A. Reece y C. Danforth, «Instagram photos reveal predictive markers of depression,» 21 06 2017. [En línea]. Available: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-017-0110-z. [Último acceso: 20 08 2019].

[74] A. S. P. M. K. M. F. Y. F. P. M. y. T. M. Pampouchidou, «Automatic Assessment of Depression Based on Visual Cues: A Systematic Review,» 2017. [En línea]. Available: https://sci-hub.tw/10.1109/TAFFC.2017.2724035. [Último acceso: 20 08 2019].

[75] M. J. K. A. &. A. H. J. Patel, «Studying depression using imaging and machine learning methods,» 2016. [En línea]. Available: https://www.sciencedirect.com/science/article/pii/S2213158215300206. [Último acceso: 20 08 2019].

[76] C. F. a. J. F. C. Y. Yang, «Detecting Depression Severity from Vocal Prosody,» 29 11 2012. [En línea]. Available: https://ieeexplore.ieee.org/document/6365169. [Último acceso: 20 08 2018].

[77] A. P. e. al, «Facial geometry and speech analysis for depression detection,» 14 09 2017. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/8037103. [Último acceso: 20 08 2019].

[78] D. J. a. H. S. L. Yang, «Integrating Deep and Shallow Models for Multi-Modal Depression Analysis — Hybrid Architectures,» 14 09 2018. [En línea]. Available: https://ieeexplore-ieee-org.ezproxy.utp.edu.co/document/8465953/metrics#metrics. [Último acceso: 20 08 2019].

[79] C. Fairbairn, Y. Yang y J. F. Cohn, «Detecting Depression Severity from Vocal Prosody,» 06 2013. [En línea]. Available: https://www.computer.org/csdl/journal/ta/2013/02/tta2013020142/13rRUxOdD6F. [Último acceso: 31 07 2019].

[80] J. Joshi, R. Goecke, A. Dhall, S. Alghowinem, M. Wagner, J. Epps, G. Parker y M. Breakspear, «Multimodal assistive technologies for depression diagnosis and monitoring,» 2013. [En línea]. Available: http://users.cecs.anu.edu.au/~adhall/Joshi_MultiModal_Depression_Analysis.pdf. [Último acceso: 31 07 2019].

[81] M. Xingchen, Y. Hongyu, C. Qiang, H. Di y W. Yunhon, «DepAudioNet: An Efficient Deep Model for Audio based Depression Classification,» 16 10 2016.

[En línea]. Available: https://dl.acm.org/citation.cfm?id=2988267. [Último acceso: 31 07 2019].

[8 2] geeksforgeeks.org, «Red neuronal en Keras: Guía practica,» 23 02 2020. [En línea]. Available: https://www.geeksforgeeks.org/confusion-matrix-machine-learning/. [Último acceso: 05 08 2020].

[8 3] USC - University of Southern California, «DAIC-WOZ Depression Database,» 2017.

[8 4] J. M. Dueñas Quesada, «Aprendizaje supervisado para la detección de amenazas web mediante clasificación basada en árboles de decisión,» 02 06 2020. [En línea]. Available: http://openaccess.uoc.edu/webapps/o2/bitstream/10609/118166/1/joseduenasTFM0620.pdf. [Último acceso: 20 06 2020].

[8 5] S. Kumar, «Everything You Need To Know About Train/Dev/Test Split — What, How and Why,» 17 03 2019. [En línea]. Available: https://medium.com/@snji.khjuria/everything-you-need-to-know-about-train-dev-test-split-what-how-and-why-6ca17ea6f35. [Último acceso: 09 12 2019].

[8 6] Pfizer Inc., «PHQ-9: Lista de verificación de nueve síntomas,» 2019. [En línea]. Available: http://www.gericareonline.net/tools/spn/depression/attachments/Dep_05_PHQ9_sp.pdf. [Último acceso: 15 01 2020].

[8 7] Sennheiser, «HSP 4-EW-3,» [En línea]. Available: https://es-mx.sennheiser.com/wireless-headworn-microphone-headmic-headset-live-performance-hsp-4. [Último acceso: 18 03 2020].

[8 8] HP, «HP Compaq 6200 Pro Small Form Factor PC Product Specifications,» 2020. [En línea]. Available: https://support.hp.com/vn-en/document/c02779493. [Último acceso: 20 07 2020].

[8 9] D. Kaspar, A. Bailey y P. Fuller, «Librosa: una biblioteca de audio de Python,» 28 05 2019. [En línea]. Available: https://medium.com/@patrickbfuller/librosa-a-python-audio-libary-60014eeaccfb. [Último acceso: 02 04 2020].

[9 0] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot y E. Duchesnay, «Scikit-learn: Machine Learning in Python,» 2011. [En línea]. Available: https://scikit-

learn.org/stable/modules/generated/sklearn.utils.resample.html. [Último acceso: 05 08 2020].

[9 1] A. Downey, «Think DSP - Digital Signal Processing in Python,» 09 08 2014. [En línea]. Available: https://greenteapress.com/thinkdsp/thinkdsp.pdf. [Último acceso: 02 06 2020].

[9 2] M. Ruiz Costa-jussá y H. Duxans Barrobés, «Diseño y análisis de filtros en procesamiento de audio,» 2012. [En línea]. Available: https://www.exabyteinformatica.com/uoc/Audio/Procesamiento_de_audio/Procesamiento_de_audio_(Modulo_2).pdf. [Último acceso: 20 01 2020].

[9 3] 123APPS, «Convertir audio - Online,» 123APPS, 2020. [En línea]. Available: https://online-audio-converter.com/es/. [Último acceso: 10 07 2020].

[9 4] L. Hernández, «DATOS NO BALANCEADOS. SOBREMUESTREO, SUBMUESTREO Y PONDERACIÓN,» 03 01 2019. [En línea]. Available: https://www.doctormetrics.com/datos-no-balanceados/. [Último acceso: 04 08 2020].

[9 5] D. Rodríguez, «El problema de desequilibrio de clases en conjuntos de datos de entrenamiento,» 04 07 2018. [En línea]. Available: https://www.analyticslane.com/2018/07/04/el-problema-de-desequilibrio-de-clases-en-conjuntos-de-datos-de-entrenamiento/. [Último acceso: 05 08 2020].

[9 6] G. Varoquaux, E. Gouillart, O. Vahtras, V. Haenel, N. P. Rougier, R. Gommers, F. Pedregosa, Z. Jędrzejewski-Szmek, P. Virtanen y C. Combelles, «Scipy Lecture Notes: One document to learn numerics, science, and data with Python,» 06 10 2015. [En línea]. Available: https://hal.inria.fr/hal-01206546/file/ScipyLectures-simple.pdf. [Último acceso: 05 08 2020].

[9 7] C. Rossant , «IPython CookBook Interactive Computing and Visualization,» 2018. [En línea]. Available: https://ipython-books.github.io/48-processing-large-numpy-arrays-with-memory-mapping/. [Último acceso: 05 08 2020].

[9 8] Keras, «Python & NumPy utilities - to_categorical function,» 2020. [En línea]. Available: https://keras.io/api/utils/python_utils/#to_categorical-function. [Último acceso: 06 08 2020].

[9 9] A. E. Ortiz, «GPU vs CPU, diferencias y similitudes,» 09 02 2019. [En línea]. Available: https://blog.hostdime.com.co/gpu-cpu-diferencias-similitudes/. [Último acceso: 09 08 2020].

[10] O. Tamara y C. Manterola, «International Journal of Morphology,» 12 19 2016. [En línea]. Available: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0717-95022017000100037#f4. [Último acceso: 22 07 2019].

[11] OMS, «Suicidio,» 02 09 2019. [En línea]. Available: https://www.who.int/es/news-room/fact-sheets/detail/suicide. [Último acceso: 17 07 2019].

[12] American Psychiatric Association, «Manual Diagnóstico y estadístico de Trastornos mentales,» 10 2018. [En línea]. Available: https://psychiatryonline.org/pb-assets/dsm/update/DSM5Update_octubre2018_es.pdf. [Último acceso: 22 07 2019].

[13] OMS, «Pocket Guide to the ICD-10 "CIE-10" Classification of Mental and Behavioural Disorders,» 1994. [En línea]. Available: https://apps.who.int/iris/bitstream/handle/10665/42326/8479034920_spa.pdf?sequence=1&isAllowed=y. [Último acceso: 22 07 2019].

[14] S. Pita Fernández y S. Pértegas Díaz, «Pruebas diagnósticas: Sensibilidad y especificidad.,» 07 12 2010. [En línea]. Available: https://www.fisterra.com/mbe/investiga/pruebas_diagnosticas/pruebas_diagnosticas.asp. [Último acceso: 22 07 2019].

[15] K. Kiefer, «pyAudioAnalysis,» 17 02 2019. [En línea]. Available: https://github.com/tyiannak/pyAudioAnalysis. [Último acceso: 18 06 2019].

[16] American Psychiatric Association, «DSM-IV - Manual diagnóstico y estadístico de los trastornos mentales,» 1995. [En línea]. Available: http://www.mdp.edu.ar/psicologia/psico/cendoc/archivos/Dsm-IV.Castellano.1995.pdf. [Último acceso: 22 07 2019].

[17] A. López Arias, «¿Qué son y por qué se hacen las pruebas de tamizaje?,» 06 05 2015. [En línea]. Available: http://uvsalud.univalle.edu.co/comunicandosalud/wp-content/uploads/2015/05/06.05.15-Qu%C3%A9-son-y-por-qu%C3%A9-se-hacen-las-pruebas-de-tamizaje.-p%C3%A1g-3.pdf. [Último acceso: 23 11 2019].

[18] Project Jupyter , «jupyter.org,» 29 07 2020. [En línea]. Available: https://jupyter.org/. [Último acceso: 02 08 2020].

[10] RADIOWORLD, «Técnicas de reducción de ruidos de audio,» 23 10 2013. [En línea]. Available: https://www.radioworld.com/global/tcnicas-de-reduccin-de-ruidos-de-audio. [Último acceso: 04 08 2020].

[11] Oromen, «Ondas, Sonido y Espectros,» 22 11 2017. [En línea]. Available: https://www.youtube.com/watch?v=s7DeLWXeWgY. [Último acceso: 05 08 2020].

[11] L. Gonzalez, «Introducción a la librería Scikit-Learn de Python,» 02 10 2018. [En línea]. Available: https://ligdigonzalez.com/libreria-scikit-learn-de-python/. [Último acceso: 05 08 2020].

[11] A. M. Manzano, «Transformada rápida de Fourier Implementación y algunas aplicaciones,» 26 06 2018. [En línea]. Available: https://www.um.es/documents/118351/9850722/Mart%C3%ADnez+Manzano+TF_48705250_v2.pdf/c44507c8-e990-4aac-b282-927acadcedd1. [Último acceso: 05 08 2020].

[11] P. Recuero, «Machine Learning a tu alcance: La matriz de confusión,» 23 01 2018. [En línea]. Available: https://empresas.blogthinkbig.com/ml-a-tu-alcance-matriz-confusion/. [Último acceso: 11 08 2020].